Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

The Timing of Signals in Skill: R. CONRAD 365

Transfer from Verbal Pretraining to Motor Performance as a Function of Motor Task Complexity: W. F. BATTIG 371

Discrimination Learning as a Function of Reversal and Nonreversal Shifts: R. T. KELLEHER 379

Some Sources of Difficulty in Solving Simple Problems: J. J. GOODNOW AND T. F. PETTIGREW 385

Effects of Stimulus Similarity on Discrimination Learning: C. C. SPIKER 393

Constant-Sum Judgments of Facial Expression: T. ENGEN AND N. LEVY 396

Perceptual Behavior as Related to Factors of Associative and Drive Strength: J. ENGLER AND J. T. FREEMAN 399

Response Generalization as a Function of Intratask Response Similarity: M. E. NOBLE AND H. P. BAHRICK 405

A Failure to Find Second-Order Semantic Generalization: I. MALTZMAN AND L. O. BROOKS 413

Task Instructions for Anagrams Following Different Task Instructions and Training: I. MALTZMAN, E. EISMAN, L. O. BROOKS, AND W. M. SMITH 418

Further Gradients of Error Reinforcement Following Repeated Rewarded Responses: M. H. MARX AND F. E. GOODSON 421

Sampled-Data Tracking: Sampling of the Operator's Output: C. A. BENNETT 429

American Psychological Association

Vol. 51 No. 6 June 1956

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1956 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1956 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 51, NO. 6 JUNE, 1956

THE TIMING OF SIGNALS IN SKILL[1]

R. CONRAD

Applied Psychology Research Unit, Cambridge, England

Some years ago Bartlett (**2, 3**) called attention to the importance of timing in skilled behavior. The term was used in its ordinary everyday sense to refer to the temporal "layout" of a series of responses. Thus good timing would show in smooth performance, bad timing in irregular, jerky performance. Although both the descriptions of good and bad timing, and the assessment of its importance were penetrating, the concept lacked precision as a scientific term.

Keeping close to the sense in which Bartlett used the term, the author (**5**) showed experimentally that in a sensorimotor skill which permitted some latitude in the moment of initiation of a response, the moment chosen tended to be that which gave *Ss* more time for response than they would have had if their responses had more exactly matched the signal series. That is, if Signal 1 was seen to be followed closely by Signal 2, then Response 1 was advanced in time. But if Signal 2 was seen to be some time after Signal 1 then Response 1 tended to be withheld until after Signal 1 had occurred. Based on these results, timing was defined (**7**) as "creating the most favorable temporal conditions for response."

The conditions in which timing occurs in any skill are defined by the nature of the particular task. The degree of constraint to which the timing must submit will vary. In unpaced tasks, constraints are slight. The moment of initiation and the duration of responses are relatively unrestricted. But in paced tasks *S* has little or no power to decide *whether* or not events requiring action should occur in the environment or display. He might have freedom to determine when to deal with them, however, and in many cases he will be free to decide *when* they should occur.

This can happen in two different ways. In the first, *S* cannot affect the time of onset of a signal. The "timetable" is laid down by conditions which are beyond his control. Skills in which the display consists of a number of dial indicators are often of this kind. These might be the control of chemical processes, or the work of aircrew when the flight char-

[1] The work reported was sponsored by the Medical Research Council, and *Ss* were made available by the Royal Navy. Miss Barbara A. Hille carried out the testing and the analysis of records, and the author is grateful to Dr. N. H. Mackworth for reading and commenting on this paper when in draft form.

acteristics are put in by another person, or those periods in many competitive games when the ball or the initiative is with the opponent. This case has been considered experimentally by the author (**5**).

In the second case, *S* can, within definable limits, change the time interval between successive signals. Usually the limiting factor is that he cannot change the relative differences between successive intervals, but he can slow down or speed up the rate of change of the environment or display as a whole. Though far from being a complete description, this type of control is an important feature of, for instance, the skill of car driving, and other skills in which the operator can vary his own "speed" relative to that of the environment. It is obvious that changing speed in this way could be a highly effective method of timing as it has been defined above.

In the present experiment, a task largely of this second kind was used. A series of signals in a complex sensorimotor task was available to *S* who had continuous control over the rate at which they occurred. The object of the study was first to observe how he utilized this control, and second to observe whether or not the quality of his performance changed when he was deprived of this control.

Method

Apparatus.—The apparatus used was the Multi-Dial Display described by Carpenter and Conrad (**4**) and Conrad (**6**). An eight-dial 4×2 display was used in this case. Each dial of this apparatus carries a visible revolving pointer, which stops at clearly indicated marks at the 6 o'clock and 12 o'clock positions if *S* makes no response before the pointer reaches either mark. When a pointer thus stops, the single hand on an 18-in. diameter clock also visible to *S* registers minus one. Further losses are registered if a pointer remains stopped, the rate of loss being proportional to the speed of the pointer when running. Pointers can be restarted by pressing the corresponding spring-loaded switch on a control panel which *S* has before him. Restarting a pointer neither gains nor loses points.

In order to gain points on the clock *S* must decide in advance which of the eight pointers will next reach the 6 o'clock or 12 o'clock mark, press the appropriate switch just before it reaches the mark, and hold it down while the pointer *crosses* the mark and so continues revolving without stopping. Each such successful response gains one point. Thus a high over-all score can be obtained by making many correct decisions accompanied by correct action with few and brief pointer stoppages.

The pointers are driven through a chain of gear wheels by a single motor. The relative pointer speeds are therefore fixed, but the speed of each is different from that of the others. The faster the pointers revolve, the more frequently will decisions be required.

A new feature now added to the Multi-Dial Display is that pointer speed is entirely under *S*'s control. He can vary the speed of the main motor by turning a potentiometer. This is sprung in such a way that the motor (and pointers) stops if he releases this control, and increases speed linearly as he turns the potentiometer in a clockwise direction. The *S* operates this speed control with his left hand, and the panel of switches with his right hand.

Because the pointers revolve at different rates, if a constant motor speed is maintained the intervals between successive signals will be unequal, and in an effectively nonrepeating sequence. Cox and Smith (**8**) have shown that the distribution of this interval approximates to that of the negative exponential. This is the basic temporal pattern of events in the display. But because the speed-control potentiometer is almost instantaneous in its effect, its use can not only change the value of the mean interval between signals, but it could be used to change individual intervals, affecting the entire nature of the distribution.

The motor speed is recorded by summing its revolutions during tests on a numeral revolution counter. A multichannel Biro pen event marker records when a pointer stops, and the moment at which pointers cross the target mark when a successful response is made. These events are recorded separately for each dial on paper moving at 1 cm./sec.

Procedure.—Eighteen naval ratings served as *S*s. On each of five days they practiced for 30 min. They began with only three dials in the display and worked up to reasonable proficiency with eight dials. The main instruction given

was to make as high a score as possible and *Ss* were told each other's score. The *Ss* quickly learned that driving the pointers too fast tended to lose them more points than they gained because many stoppages occurred. Driving the pointers too slowly, on the other hand, resulted in very few stoppages, but scored points only slowly. The task itself is intrinsically interesting and the motivation of *Ss* was exceptionally high.

During these practice days, the sessions were broken up into short periods of 5- or 10-min. duration. On Day 6 a 10-min. experimental test run was made. At the end of it, the clock score was noted, and the average speed in revolutions per minute at which the motor had been driven was calculated from the figure showing on the revolution counter. The speed-control potentiometer was then switched out of the system. The *Ss* were told that the motor would be set automatically to run at a constant speed, exactly the average of the previous test run. Under this condition they would have no control over the distribution of the interval between signals, and in fact all their attention could be given to the problem of preventing pointers from stopping, again using only the right hand. The general instruction was to try to beat their score of the previous test. This condition was not new as it had been frequently introduced during the practice days. The appropriate speed was set, which was different for each *S*, who worked for another 10 minutes.

It may appear odd always to present the two conditions in the same order to each *S*. In fact, there is no other possibility in the present case. The paced speed must be identical to that chosen by *S* in the self-paced condition, and must therefore succeed it. This constraint would tend to add reliability to the results since there would be a slight practice effect in the second condition.

Results

The effects of timing on the display.—The definition of timing adopted implies an active relationship between *S* and display. More specifically, the definition implies that *S* modifies the temporal structure of the task so as to make it easier. It is postulated that ideal timing would tend to aim at changing the time structure of the signal series from whatever it inherently was to a series of equally spaced events; this postulation assumes that the events are equivalent.

There are two main reasons why in the present case, and in most other cases, the ideal aim is unlikely to be achieved. First, because of the known variability of human performance, and second, because with several concurrent signal sources comprising the display there will be a number of occasions when two or more sources present signals exactly simultaneously. It has been indicated that with this apparatus a zero interval cannot be made into anything else. In practice this restriction will extend to all very short intervals even though they are not infinitely short, so that a less ideal but more likely temporal structure is one in which the interval between events is normally distributed. Finally, as a working principle it is suggested that the main effect of timing on an equivalent-signal series would be to reduce the variability of interval size.

The measure of variability adopted was a log transformation of *SD*, the data coming from the measurement of the interval between every signal in the 10-min. test for each *S*. For the self-paced condition, the mean of the *SD*'s of the 18 *Ss* was .362.

The significance of this value can only be assessed if it is known what the value would have been if *Ss* had been unable to exercise control over the signal series. This value is provided by the second condition, since for each *S*, the *mean size of interval* was the same as for the first. For the second test the mean *SD* was .434. Every *S* had a lower *SD* for the self-paced than for the paced condition. A t test showed that the difference was significant at better than the .001 level.

The histograms in Fig. 1 compare some typical distributions of the

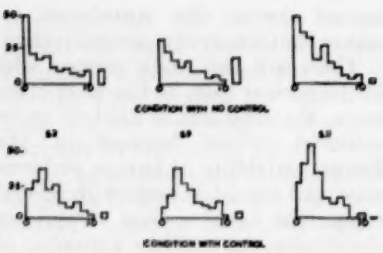

FIG. 1. Examples of distributions of the interval between signals under the two test conditions. (Distributions are in steps of 1.0 sec. Final step includes all intervals greater than 10.0 sec.)

interval between signals for the two test conditions. For clarity these are shown in the measured time units rather than in logarithmic units. It is clear that the principal effect is that *Ss* try to lengthen very short intervals, and to a lesser extent to shorten very long time intervals. From the point of view of the aim of the skill, there are sound reasons for this. For instance, changing an interval of say, .5 sec., into one of 2.0 sec. means almost certain success with two consecutive responses, when the probability of two failures was high and of one failure almost certain.

By reducing the duration of the long intervals that would occur, *S* allows himself more opportunity to achieve a good score by "feeding" himself, in the finite test period, more signals that are likely to lead to good response.

The effect of timing on the quality of performance.—With the equipment described, responses can be clearly classified into successful and unsuccessful, or preferably, positive and negative responses. Positive responses are those which prevent a pointer from stopping, negative responses are those which merely restart stopped pointers. The numerical difference per unit time between positive and negative responses is the rate at which *S* advances his score, i.e., a measure of success in dealing with the display. Performance under the two test conditions can be compared by using this measure. For the first condition when they could modify the temporal structure of the signal series, the mean score for the 18 *Ss* was 98. For the second condition (i.e., the same signal speed) the mean score was 83. The difference between these means is significant at better than the .001 level.

There is therefore considerable advantage to be gained as a result of timing insofar as it reduces temporal variability within the signal series. This being the case, one would expect that the more skillful the timing, i.e., the larger the effect on the structure of the signal series, the greater should

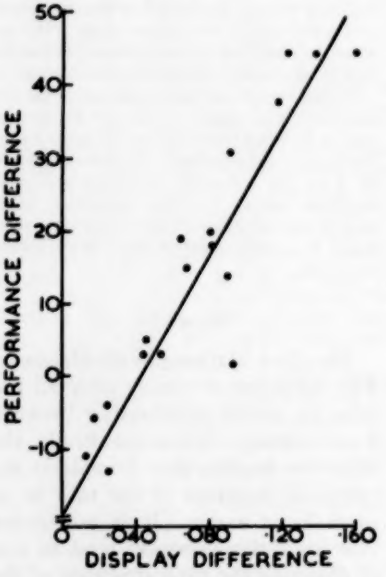

FIG. 2. Relationship between change in the temporal structure of the signal series and change in performance.

be the beneficial effect on performance.

For each *S* we have a measure of interval variability under the two conditions, the difference representing the effect on the display of *S*'s control. We also have measures of performance success for the two conditions. One would expect a close positive association between differences in interval variability and differences in performance. The most convenient index of performance is that referred to—the numerical difference between positive and negative responses. The larger this difference is, the better, in terms of the aim of the task, is performance.

The relationship is shown in Fig. 2. It is evident that *S*s who make large differences to the display as a result of timing, achieve great improvement in their performance scores. The correlation between these two variables is .92.

Discussion

It is clear from these results, not only that timing is an important component of skilled performance, but that it is important insofar as it enables *S*s to re-allocate time within the framework of the task. An *S* whose timing is bad is compelled continually to make the best he can of conditions imposed by the display. But good timing forces the display into a temporal form which is convenient for *S*. This form has been shown to be in the direction of a normal distribution of the interval between successive signals.

Throughout this paper, the term *timing* has referred to actual modifications of the display. This has been an arbitrary restriction of the term, since it has been previously demonstrated (5) that timing in the sense defined is equally a component of skills in which the nominal display cannot be affected by *S*. In such cases it was shown that *S* advances or retards response according to his assessment of objective and fixed display demands. When, as in the present case, he can change the display, timing of the discrete key responses is not abandoned. Hurried jerky responses can still be distinguished from the smoother apparently languid movements from key to key of the more skilled *S*. In the present case the definition and measurement of well- and badly-timed key responses would be technically difficult. The intention here was to examine the role of the timing of signals when continuous change of display and control were characteristic, and whether the cited definition of timing maintained its value.

In discussing the mathematical nature of the changes effected by timing, the essential condition of equivalence of signal cannot be overstressed. One would not expect any equally spaced series of signals to be optimal if some of the signals were inherently more difficult to deal with. This is without doubt the reason why Knowles, Garvey, and Newlin (9) were able to report an advantage for performance in their self-paced compared with their paced condition. In fact, when *S*s could manipulate the temporal structure of a series of signals of two levels of difficulty, one might expect that a bimodal distribution of the interval between them would result.

Finally, the experiment reported is relevant to the discussion of the interaction of two separate components of a motor skill. Classical studies and more recent ones (1) have emphasized the role of a second component as interfering. The present study illustrates a case where the second component might interfere with the first especially during the early part of learning. But high levels of performance are only possible if the second component is integrated into the whole, when it becomes of positive assistance. This is merely to make the important distinction between what is truly a *component* of a motor skill and must logically contribute towards its aim, and purely extraneous, separate, distracting tasks which can

play no part at all in the furtherance of another skill.

Summary

In terms of a definition of "timing," it is postulated that in certain skills the effect of timing will be to change the temporal structure of an approximately random series of signals to one in which the interval between signals is more normally distributed, and that the effect of this change on performance is advantageous.

The experiment employed a complex sensorimotor task in which a multidial display presented signals for response at approximately random time intervals. The 18 male *S*s were tested under two conditions. In the first they could operate a control by means of which the interval between each pair of signals could be changed. In the second condition they were deprived of this control. It was arranged that the display would present signals at a rate exactly equal to that which each *S* had selected in the first test, but the inherent temporal structure of the signal series (approximately random) was left undisturbed.

The two resulting distributions of the interval between signals were compared. It is shown that the inherent distribution of the second test was changed in the first test in the direction of normality, and that the *SD* (log time) was reduced by every *S*. This change had the effect of significantly improving performance, and a close positive association is demonstrated between the extent of the difference between the two *SD*'s and the extent of the improvement in performance score.

Drawing on these results, the role of timing in skill is discussed, a distinction being drawn between the timing of continuously graded and intermittent response.

References

1. Baker, K. E., Wylie, R. C., & Gagné, R. M. The effects of an interfering task on the learning of a complex motor skill. *J. exp. Psychol.*, 1951, **41**, 1–9.
2. Bartlett, F. C. Fatigue following highly skilled work. *Proc. Roy. Soc.*, 1943, **B 131**, 247–257.
3. Bartlett, F. C. The measurement of human skill. *Brit. med. J.*, 1947, **4510**, 835–838; **4511**, 877–880.
4. Carpenter, A., & Conrad, R. An apparatus providing signals at "random" intervals. *Quart J. exp. Psychol.*, 1953, **5**, 87.
5. Conrad, R. Adaptation to time in a sensorimotor skill. *J. exp. Psychol.*, 1955, **49**, 115–121.
6. Conrad, R. Some effects on performance of changes in perceptual load. *J. exp. Psychol.*, 1955, **49**, 313–322.
7. Conrad, R. Timing. *Occup. Psychol.*, 1955, **29**, 3.
8. Cox, D. R., & Smith, W. L. The superposition of several strictly periodic sequences of events. *Biometrika*, 1953, **40**, 1–11.
9. Knowles, W. B., Garvey, W. D., & Newlin, E. P. The effect of speed and load on display-control relationships. *J. exp. Psychol.*, 1953, **46**, 65–75.

(Received June 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

TRANSFER FROM VERBAL PRETRAINING TO MOTOR PERFORMANCE AS A FUNCTION OF MOTOR TASK COMPLEXITY [1]

WILLIAM F. BATTIG

University of Wisconsin [2]

Several recent experiments have been concerned with the effect of verbal pretraining on subsequent performance on a motor task. Since the results of such studies have not been consistent, a next logical step in research is to try to identify the variables accounting for this inconsistency.

The purpose of this experiment was to determine the effect of the complexity of the motor task on the nature and amount of transfer from verbal pretraining. An examination of previous studies indicates that this variable may account for much of the inconsistency in results. Performance on simple discrimination-type motor tasks, such as pressing a switch (**1, 7, 9, 11, 13**) or moving a lever (**5, 14, 16**), is usually facilitated by verbal pretraining. However, with more complex tasks such as accurate lever-positioning (**2**) and tracking (**4**), no evidence for positive transfer from verbal pretraining has been obtained.

The nature of the verbal pretraining conditions may also be an important variable. McAllister (**14**) found that verbal pretraining in which both the stimuli and responses are relevant to the motor task (*relevant S–R*) produces significant positive transfer to motor performance, but verbal pretraining in which the stimuli are relevant but the responses are irrelevant (*relevant S*) shows no positive transfer. The terms relevant S–R and relevant S correspond to "verbal pretraining" and "stimulus predifferentiation" respectively, which have been used by other investigators (**1, 7, 11**). The present study also provides a check on the effect of this variable.

Method

Apparatus.—The finger-positioning apparatus used in the present experiment consisted of three relatively distinct units: (*a*) response unit, (*b*) display panel, and (*c*) control unit. The apparatus, from *S*'s point of view, is shown in Fig. 1.

The response unit consisted of four pairs of Mallory 7223C three-position spring-return lever-action switches mounted on a rectangular metal cabinet. The two switches of a pair were mounted in the vertical dimension, each pair being separated by 1¼ in., with a 3-in. separation between the middle two pairs. A Lucite key (¾ × 1 × 3 in.) covered by black rubber tape was attached to each switch. The switches were operated by placing the finger between the two keys of a pair and moving the finger either up or down. The upper key of a pair had the lowest of the three positions as its normal resting position, and could be moved only upward, automatically returning to the normal position when released. Similarly, the lower key was normally in its highest position and could be moved only downward. As only one key of a pair could be moved at a given time, there were five possible positions for each pair, one with both keys in the normal position, two of the upper key, and two of the lower key. The two keys of a pair were 1 in. apart, parallel to each other and to the table when in the normal

[1] This research was carried out while the author was a National Science Foundation predoctoral fellow, and is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Wisconsin. The author wishes to express his appreciation to Dr. W. J. Brogden, under whose guidance the research was done, for his many helpful suggestions.

[2] Now a National Science Foundation postdoctoral fellow at Stanford University.

FIG. 1. The apparatus from *S*'s point of view.

position. There was a 60° angle between successive key positions; however, the linear extent of movement could vary from ¾ in. to 1¾ in., depending on where *S*'s fingertip was placed on the key.

The display panel consisted of four double columns of 6-w. lamps, five in each column. The lights in a column were 1¼ in. apart, and the two columns in a double column were separated by 1 in., with 2 in. between each double column. The column at the left of each double column was colored red, that at the right was colored green. Each column of green lights was connected with one pair of response keys. The five green lights from top to bottom corresponded directly to the five key positions from top to bottom.

The control unit was operated by *E* and consisted of switches for controlling the red lights on the display panel, plus the appropriate controls for scoring and turning on and off various components of the apparatus.

The stimulus for the motor task consisted of the lighting of one red light from each column. The *S*'s task was to light simultaneously each green light corresponding to the red lights which were lit as quickly as possible by moving the keys to the appropriate positions with the fingers. This constituted a *match*. Self-paced conditions were used so that after a match had been made and held for approximately ½ sec., a new pattern of red lights was presented, and *S* immediately proceeded to match the new pattern. Performance was scored in terms of number of matches in a 1-min. trial.

The response unit and display panel were placed at opposite ends of a 6-ft. table. The control unit was on a table behind a partition to *S*'s right, at which *E* sat. The *S* sat in an adjustable chair, so that his elbow height was approximately that of the response keys.

Subjects.—The *S*s were 240 University of Wisconsin undergraduate psychology students, 120 of each sex, who received class points for their services. They were unsystematically assigned to 12 groups of 20 *S*s each, with an equal number of males and females in each group.

Design.—The 12 groups formed a 3×4 factorial design, with three types of pretraining conditions and four levels of motor task complexity, differentiated in terms of the number of fingers used on the task in four steps from one to four. The one-finger groups used only the index finger of the right hand on the pair of keys and double column of lights at the extreme right of the apparatus. The two-finger groups used the index and third fingers of the right hand on the two key-light units at the right. The three-finger groups used in addition the index finger

TABLE 1

LETTERS USED IN RELEVANT S PRETRAINING

Column			
1	2	3	4
Z	O	N	T
Q	A	R	F
X	U	P	B
Y	I	M	K
V	E	L	S

of the left hand, operating all key-light units except the one at the extreme left. The four-finger groups used the index and third fingers of both hands on all four key-light units.

The pretraining conditions were as follows: (*a*) *Control:* No pretraining was given.[3] (*b*) *Relevant S:* The green light columns were covered by cards containing letters of the alphabet, one letter corresponding to each red light. The *S*'s task was to pronounce aloud nonsense words formed from the letters reading from left to right which corresponded to the red lights of the stimulus patterns. The letters for the four-finger complexity are shown in Table 1. For each successively lower level of complexity the column of letters at the right was eliminated, and the other columns moved one position to the right. The second letter in each word was always a vowel, making the words pronounceable yet relatively meaningless. Nothing was said about memorizing either the words or letters. (*c*) *Relevant S-R:* The green lights were turned off, and *S*'s task was to describe verbally the key positions which would match the red light patterns, using the terms two-up, one-up, center, one-down, or two-down.

Procedure.—The experimental session was about 50 min. in duration. Upon entering the experimental room, all *S*s were seated at the apparatus and given standard instructions regarding the operation of the apparatus and procedure to be followed. The control groups proceeded immediately to practice on the motor task. The pretraining given the relevant S and relevant S–R groups consisted of the presentation of 50 red light patterns at 10-sec. intervals, to which the appropriate verbal response was made. The relevant S groups were then asked to write down as many of the letters as they could remember (this was also done at the close of the experiment), in order to provide a check on any relationship between learning of the letters and motor performance. Following the pretraining, a summary of the initial instructions was given, along with the suggestion that use should be made of the verbal pretraining, since it should be helpful.

All groups performed under identical conditions on the motor task, receiving 20 1-min. trials with a 15-sec. rest after each trial, and a 1-min. rest after every fifth trial. After each trial *S*s were told their scores in terms of number of matches.

A single sequence of 50 stimulus patterns was constructed for each level of complexity and used repeatedly throughout motor practice. The 50 pretraining patterns also followed this sequence. For each higher level of complexity this sequence was identical to that for the next lower level, except for the addition of another red light to each pattern. This made the sequence identical for all groups on the key-light unit at the right of the apparatus, and so on. The sequence was constructed on the basis of criteria designed to provide that each possible combination of finger positions appeared with approximately equal frequency, and to minimize sequential dependencies. Fifty different patterns were used for the three- and four-finger complexities, but only five were used with one-finger, and 24 for two-finger complexities (the pattern with both lights in the center position was not used).

RESULTS

Since analysis of variance revealed no significant sex difference ($F < 1$), the two sexes were pooled in analysis of the results. The mean number of matches for each group on each trial is presented in Fig. 2.

An application of Bartlett's test (**8**) on the total match scores gave a χ^2 of 39.50 ($p < .001$), causing rejection of the hypothesis of homogeneity of variance. The source of heterogeneity was such that no rational transformation was applicable. To compensate for the increase in Type I errors which may result from this violation of the assumptions underlying the analysis of variance,

[3] The control groups were run prior to the verbal pretraining groups and were given 30 trials of motor practice. Only the data from the first 20 trials are presented here. The additional data and further analyses bearing on complexity of the motor task, along with a justification for the use of the matches score, can be found in the dissertation on which this report is based (**3**).

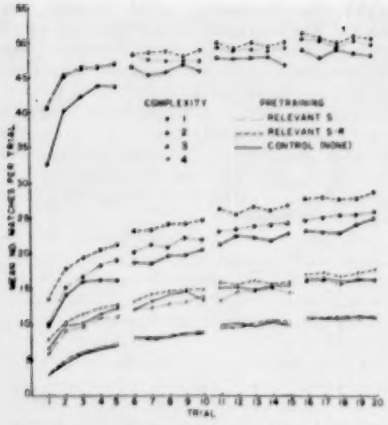

FIG. 2. Mean number of matches per trial (matches score) for each of 12 groups. Breaks in curves signify longer 1-min. rest interval.

the 1% level of significance was adopted instead of the normal 5% level.

All the data of Fig. 2 were included in a single analysis of variance, presented in Table 2, which showed all main effects and interactions to be significant. The term of primary interest is the complexity by pretraining (C × P) interaction, which indicates that the effect of the pretraining conditions varies with the complexity of the motor task. Figure 2 shows the source of this interaction, verbal pretraining was facilitative for the one- and two-finger, but not for the three- and four-finger complexities.

Separate analyses of variance on the three pretraining conditions within each level of complexity were carried out to further specify the nature of this interaction. Pretraining was significant for one finger ($F = 6.63$) and two fingers ($F = 7.70$), but not for three fingers ($F = 2.14$) and four fingers ($F = 0.19$). Duncan range tests (6) showed that both verbal pretraining conditions were superior to the control group for one finger, but there was no significant difference between relevant S and relevant S–R pretraining. For two fingers, relevant S–R pretraining was significantly superior to both relevant S and control conditions, which did not differ significantly from each other.

As is obvious from Fig. 2, there is a significant difference in motor performance between levels of complexity, accounting for the significant complexity (C) term in Table 2. Duncan range tests showed that each level of complexity differed significantly from each other level. The significant pretraining (P) term indicates a difference in the over-all effects of the pretraining conditions in this experiment. Relevant S–R pretraining was significantly superior to both relevant S and control conditions, but relevant S pretraining did not differ from the control groups, according to Duncan range tests.

The remaining terms of Table 2 are concerned with learning effects. The significant trials (T) term simply shows that there was over-all improvement in performance with practice, while the C × T, P × T, and C × P × T interactions indicate that

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF MATCH SCORES

Source	*df*	Mean Square	*F*
Complexity (C)	3	354131.5	2705.36*
Pretraining (P)	2	1622.1	12.39*
C × P	6	447.0	3.41*
Ss/Groups	228	130.9	—
Trials (T)	19	1974.3	731.22*
C × T	57	41.4	15.33*
P × T	38	5.8	2.15*
C × P × T	114	6.7	2.48*
Residual	4332	2.7	

* Significant at .01 level.

the course of learning varied as a function of both complexity of the motor task and type of pretraining, and that the relationship between course of learning and pretraining varied with the level of complexity. These interactions are largely attributable to the fact that the one-finger task was quite simple, so that learning and approach to asymptotic performance was more rapid than for the other three complexities. Analysis revealed a significant P × T interaction only for the one-finger complexity, indicating that differences between pretraining conditions diminish with practice for this complexity, but not for the three more difficult levels of complexity.

To further specify the nature of these variations with practice, separate analyses were carried out on the first and last trials. The C × P interaction was significant for Trial 1 ($F = 11.63$), but not for Trial 20 ($F = 1.67$), indicating that the obtained variation in the effect of verbal pretraining with complexity may be specific to the early stages of practice.

There is some question as to the equality of the units of measurement for the four levels of complexity, since a single match represents a progressively greater amount of performance as complexity increases. To adjust for this difference, a percentage transfer measure was computed for each verbal pretraining group, using a formula suggested by Gagné, Foster, and Crowley (**12**).

$$\text{Per Cent Transfer} = \frac{T - C_1}{C_f - C_1}$$

In this formula, T is the score of the verbal pretraining group, C_1 is the score of the control group at the corresponding stage of practice, and C_f is the score of the control group on its last trial. These scores were computed for several stages of practice, with similar results, and the mean of the first five trials was selected for presentation as the best estimate of initial transfer. It should be noted that the validity of these scores is limited by the absence of a measure of the highest possible score on the motor task. The final control group scores were selected as the best available estimate.

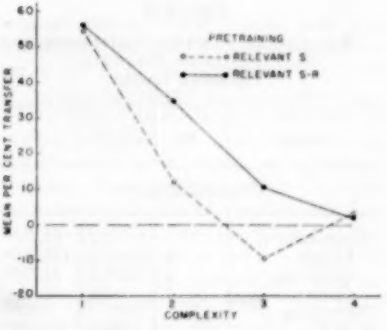

FIG. 3. Percentage transfer scores for each group given verbal pretraining. Complexity in terms of number of fingers used on the motor task is on the abscissa.

The percentage transfer scores for relevant S and relevant S-R pretraining for each level of complexity are presented in Fig. 3. Both curves show a consistent decrease in percentage transfer as complexity increases, the only reversal coming between three- and four-finger complexities in the relevant S condition.

Bartlett's test gave a χ^2 of 8.95 ($P > .05$), so that the hypothesis of homogeneity of variance is tenable for the percentage transfer scores, so the normal 5% level of significance was used. These data were analyzed by means of orthogonal polynomials (**10**), summarized in Table 3. In the combined analysis, the average of the two curves of Fig. 3 is shown to have significant linear and quadratic com-

TABLE 3

ANALYSES OF VARIANCE WITH ORTHOGONAL POLYNOMIALS OF PERCENTAGE TRANSFER SCORES

Source	*df*	Mean Square	*F*
Combined Analysis			
Complexity (C)	3	25788.7	52.10**
Linear	1	65238.1	131.79**
Quadratic	1	11575.3	23.38**
Cubic	1	552.6	1.12
Pretraining (P)	1	4779.7	9.70**
C × P	3	1508.1	3.05*
Linear	1	61.2	0.12
Quadratic	1	4448.9	8.99**
Cubic	1	14.0	0.03
Error	152	495.0	
Relevant S			
Complexity	3	15345.0	18.19**
Linear	1	30651.3	36.33**
Quadratic	1	15188.3	18.00**
Cubic	1	195.3	0.23
Error	76	843.8	
Relevant S–R			
Complexity	3	11951.8	13.94**
Linear	1	34648.1	40.41**
Quadratic	1	835.9	0.97
Cubic	1	371.3	0.43
Error	76	857.5	

* Significant at .05 level.
** Significant at .01 level.

ponents. The significance of the C × P interaction shows that the shapes of the two curves differ significantly. Only the quadratic component of this interaction is significant, indicating that this shape difference lies in the quadratic component. Analysis of the two curves separately reveals that the relevant S–R curve contains only a significant linear component, but the relevant S curve has both linear and quadratic components. The reversal between points 3 and 4 on the relevant S curve is not significant by Duncan range test.

In the relevant S pretraining, *Ss* were asked to write down the letters used during pretraining on two occasions. These responses were scored numerically and correlated with motor performance, yielding correlations ranging from .44 to − .11 for the various groups. This provides little or no evidence for memorization of these letters as a factor in motor performance. An interesting sidelight is the fact that for three of the four relevant S groups, these memory scores were higher after motor practice than immediately after verbal pretraining, although this difference is not significant.

DISCUSSION

Analysis of the data clearly demonstrates that the amount of transfer from verbal pretraining decreases as complexity of the motor task increases, according to the present dimension of complexity. A suggested explanation is based on a distinction between two processes involved in any motor task, called *discovery* and *performance* by Melton (**15**). The discovery process involves obtaining some sort of information, usually through perception of stimuli, and the selection of the motor responses to be made on the basis of this information. The performance process involves the actual execution of the motor responses. If it is assumed that verbal pretraining affects primarily discovery, and has little or no effect on performance, positive transfer from verbal pretraining should increase as the relative importance of discovery increases. Results of previous studies are consistent with this hypothesis, as positive transfer has usually been shown in simple discrimination tasks in which discovery is of paramount importance (**1, 7, 9, 11, 13, 14, 16**), but not with lever-positioning (**2**) and tracking (**4**), in which performance is more important.

In the present study, complexity of both discovery and performance was

varied, but it is probable that the relative importance of performance became greater with increased complexity. If this is so, the obtained results are in line with the above hypothesis. However, it remains to be explained why *no* positive transfer was found for three- and four-finger complexities. In absolute terms, there is more discovery involved as complexity increases, so if verbal pretraining facilitates discovery, some positive transfer should have resulted for the more complex conditions. To completely account for the results, it must be hypothesized in addition that positive transfer from verbal pretraining decreases with increased discovery complexity. This can readily be tested experimentally by increasing the number of S–R pairs on the Gagné discrimination task (**11**) or the Star Discrimeter (**14**), which would increase discovery complexity while holding performance complexity constant.

The results can also be explained by assuming an upper limit on the number of things that an individual can react to differentially at a given time, and that increased motor-task complexity in any dimension increases the number of stimulus and/or response elements involved in performance. Presumably the beneficial effect of verbal pretraining lies in the addition of new cues relevant to the motor task. In performing a simple task, such cues can be utilized without disruption of other factors. With a complex task, however, utilization of these added cues will interfere with other essential elements of the task, producing a decrement in performance. Under these conditions, *S* will perform best by ignoring the verbal cues, so that verbal pretraining would have no effect.

The significant over-all superiority of relevant S–R to relevant S pretraining, with no difference between relevant S and no pretraining, supports the findings of McAllister (**14**). A significant difference in the form of the functions relating percentage transfer to complexity was found between relevant S and relevant S–R pretraining in the present study. This result was not anticipated, and has no apparent meaning in terms of the operations involved in the two verbal training procedures.

The verbal pretraining procedures of this experiment did not require any paired-associate learning of stimulus-response relationships, unlike most of the procedures used in previous studies. The relevant S groups merely read their verbal responses, while the relevant S–R groups had them established by the instructions, so that they responded without error from the outset of verbal practice. The low correlations between motor performance and the extent to which the relevant S groups memorized the letters used during pretraining suggests that verbal learning was not an important determinant of transfer. It might be argued that the superiority of the relevant S–R groups to the relevant S groups is attributable to greater paired-associate learning under the former condition. However, there is no direct evidence for this, and since paired-associate learning was not necessary in either procedure, it is unlikely that the obtained difference can be accounted for in these terms. "Learning how to learn" and "warm-up" have been suggested as important determinants of the effect of verbal training (**7, 9**), and are probably involved here. However, the significant superiority of relevant S–R to relevant S pretraining indicates that practice of a relevant verbal response can be facilitative beyond any warm-up effects, even when little verbal learning is involved.

Whether the present results would be confirmed with other dimensions of complexity is a subject for future research. If so, the results would have application to various motor skill training situations in which "on the job training" is impractical and some substitute method of training is sought. It is suggested that verbal training of the type used here may be of considerable value when the motor task is relatively easy, but that the value of such training would diminish as task complexity is increased.

Summary

The effect of complexity of a motor task on the amount of transfer from verbal pretraining was studied. Twelve groups of 20 *Ss* each formed a 3 × 4 factorial design, with four levels of complexity in terms of number of fingers used on a finger-positioning task (one to four), and three pretraining conditions (relevant S, relevant S-R, and control). Relevant S pretraining required *S* to pronounce nonsense words formed from letters corresponding to the stimulus lights of the motor task. In relevant S-R pretraining, *S* described verbally the correct finger positions. Control groups had no pretraining. All groups then received 20 1-min. trials on the motor task.

Analysis of variance showed a significant complexity by pretraining interaction, owing to the significant superiority of both types of pretraining to the control group for one finger, and of relevant S-R pretraining for two fingers, while pretraining had no significant effect for three and four fingers. A percentage transfer measure showed a consistent and significant decrease with increased complexity for both verbal pretraining conditions. Relevant S-R pretraining showed a significant over-all superiority to both relevant S and control conditions, which did not differ from each other.

It is concluded that the amount of positive transfer from verbal pretraining to motor performance shows a consistent decrease as motor task complexity increases in terms of number of fingers used on a finger-positioning task. Possible explanations of the results are presented and discussed.

References

1. Baker, K. E., & Wylie, R. C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, **40**, 632–638.
2. Battig, W. F. The effect of kinesthetic, verbal, and visual cues on the acquisition of a lever-positioning skill. *J. exp. Psychol.*, 1954, **47**, 371–380.
3. Battig, W. F. Transfer from verbal pretraining to motor performance as a function of motor task complexity. Unpublished doctor's dissertation, Univ. of Wisconsin, 1955.
4. Battig, W. F., Gregg, L. W., Hornseth, J. P., Nagel, E. H., & Brogden, W. J. The effect on sighting proficiency of instructed practice on stimulus and response components of sighting. Unpublished manuscript, Univ. of Wisconsin, 1952.
5. Cantor, J. H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *J. exp. Psychol.*, 1955, **50**, 180–184.
6. Duncan, R. B. A significance test for differences between rank treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2** (new series), 171–189.
7. Dysinger, D. W. An investigation of stimulus pre-differentiation in a choice discrimination problem. Unpublished doctor's dissertation, State Univ. of Iowa, 1951.
8. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
9. Farber, I. E., & Murfin, F. L. Performance set as a factor in transfer of training. Paper read at Midwest. Psychol. Ass., Chicago, April, 1951.
10. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* (2nd Ed.) Edinburgh: Oliver & Boyd, 1943.
11. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
12. Gagné, R. M., Foster, H., & Crowley, M. E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97–130.
13. Goss, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46**, 419–428.
14. McAllister, D. E. The effects of various kinds of relevant verbal pre-training on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329–336.
15. Melton, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* (Rev. Ed.) New York: Macmillan, 1950. Pp. 668–690.
16. Rossman, I. L., & Goss, A. E. The acquired distinctiveness of cues: The role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.

(Received June 9, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

DISCRIMINATION LEARNING AS A FUNCTION OF REVERSAL AND NONREVERSAL SHIFTS[1]

ROGER T. KELLEHER

New York University[2]

There are two types of theoretical formulation which are applicable to the learning of consecutive simultaneous discriminations in which the stimuli vary along several dimensions. Those which assume a direct association of physical stimuli and overt responses will be referred to as single-unit S-R theories (e.g., Spence's continuity theory). On the other hand, formulations which assume the development of implicit or mediating responses with their accompanying response-produced cues will be referred to as sequential S-R theories (e.g., Lawrence's acquired distinctiveness of cues). The present study will assess the relative merits of the sequential and single-unit S-R theories by determining the effects of reversal and nonreversal shifts on discrimination learning in the white rat.

In the reversal shift, the reward values of the relevant cues during training (first discrimination) are reversed during testing (second discrimination). In the nonreversal shift, one stimulus dimension is relevant during training and another dimension is relevant during testing. It should be noted that the response of approaching the positive stimulus of the training discrimination receives adventitious partial reinforcement following a nonreversal shift. The single-unit theory (**9**) is not sufficiently developed to predict which type of shift should be accomplished more readily under these conditions; however, it does imply that both should yield negative transfer. If the partial reinforcement were eliminated, the single-unit S-R theory indicates that the nonreversal shift would be more rapid, and that only the reversal shift would yield negative transfer.

According to the sequential S-R theory, the differential reinforcement of the relevant cues during training generates "mediating responses" which give rise to a "modified stimulus pattern."[3] It is assumed that new instrumental behavior is more readily associated with this "modified stimulus pattern." Thus the reversal shift, which involves the same relevant cues on each discrimination, should lead to positive transfer. This prediction clearly follows from Lawrence's analysis of his "reversal test" (**5**, p. 185). Conversely, the nonreversal shift, which involves different relevant cues on each discrimination, should lead to negative transfer. It might be argued that the use of the same instrumental response for both training and testing violates one of Lawrence's essential conditions; however, the second and third stages of his second experiment (**5**) both involve

[1] This paper is based on a dissertation submitted to the faculty of the Graduate School of New York University. The author wishes to express his gratitude to Professor Howard Kendler to whom he is indebted for advice and encouragement in the planning and execution of this investigation.

[2] Now at the Yerkes Laboratories of Primate Biology.

[3] These responses should not be confused with receptor orienting responses (**10**) which have been manipulated in other situations (**2**, **11**). Lawrence (**4**, p. 782) conceives of these mediating responses and stimuli as theoretical constructs.

simultaneous discriminations. The only critical condition for the establishment of mediating responses is the differential reinforcement of the relevant cues over an adequate training period.

The first experiment of the present investigation will be a straightforward comparison of reversal and nonreversal shifts. The second experiment will embody a procedure which eliminates the partial reinforcement following a nonreversal shift. Both experiments include control groups which will make it possible to determine direction of transfer.

Experiment I

The plan of this experiment is to train two experimental groups on a simultaneous discrimination which involves two stimulus dimensions. After each *S* has learned the training discrimination, it is shifted to a simultaneous test discrimination involving the same stimulus dimensions. For Group R, the relevant training cues are still relevant, but the rewarded test response is the opposite of the rewarded training response. For Group NR, the irrelevant training cues become the relevant test cues. Animals of Group C are trained on a different simultaneous discrimination involving a single stimulus dimension, but are tested on the same discriminations as the experimental groups.

The sequential S–R theory predicts that Group R will show positive transfer and Group NR will show negative transfer. The single-unit theory is not sufficiently developed to predict which type of shift should be accomplished more readily. This theory does predict that negative transfer will result from both types of shift.

Method

Subjects.—The *S*s were 80 experimentally naive, male albino rats of the Wistar strain. At the beginning of the experiment they were 80–120 days of age. Eight *S*s were eliminated. Three could not learn the first discrimination, three would not perform, and two were lost through *E*'s error.

Apparatus.—The discrimination apparatus was constructed of ¼-in. plywood. The stem was 8 in. long, 4 in. wide, and 5 in. high. Its floor was 1 in. higher than that of the stimulus alleys. The entire stem was painted a flat gray. The *S* had to step over a 4½-in. air gap, which could be narrowed for adaptation training, to a 1 × 2-in. gray strip centered in front of two adjacent stimulus alleys. These alleys were each 15 in. long, 3½ in. wide, and 6 in. high. Two stimulus boxes were constructed in such a manner that black and white alleys could be presented in a right-left or left-right order. The presence or absence of chain curtains in either alley served as a second stimulus dimension (cf. **5**). Two other stimulus boxes contained gray alleys and unpainted alleys, respectively. A 1-in. hurdle constructed of ¼-in. plywood could be placed 2 in. from the entrance of either of the unpainted alleys. Behind a wooden baffle at the end of each stimulus alley was a small block of wood with a depression in the top surface for food pellets. The entire apparatus was covered by clear glass and indirectly illuminated by a 75-w. bulb located 36 in. above the air gap.

Procedure.—A feeding schedule was instituted 1 wk. before the start of the experiment and was maintained throughout. All *S*s had access to a wet mash for 1 hr. per day, and this was presented 15–20 min. after the completion of the daily trials. On Day 1 of the experiment, all *S*s were allowed 7–10 min. to explore the stem and gray alleys with the air gap closed. Five .065-gm. pellets were present in each food block during exploration. On Days 2 and 3, *S*s received eight forced trials to each of the possible combinations of black-white and chains-no chains in each spatial position. Each trial was rewarded by a single food pellet. The air gap was gradually increased to its full width during this pretraining. If *S* had not left the stem by the end of a 2-min. period, it was forced by moving the rear wall of the stem forward.

For training, *S*s were randomly assigned to one of three groups. The 24 *S*s of the control group learned the hurdle discrimination in the unpainted alleys. Of the experimental *S*s, 24 learned the black-white discrimination (with chains-no chains irrelevant), and 24 learned the chains-no chains discrimination (with

black-white irrelevant). Each cue appeared an equal number of times on the right and on the left, and no combination of two cues, such as chains and black, appeared together on more than two consecutive trials. On every day of training and testing, each *S* received 10 non-correction trials with at least 5 min. between trials. The criterion of learning was 18 correct out of 20 consecutive trials. Animals that did not learn the first discrimination in 30 days (300 trials) were eliminated.

For the test trials, each of the experimental training groups was divided into Groups R and NR. These groups were equated for speed of learning on the training discrimination. Equal numbers of *S*s from Group C were assigned to each of the possible positive test stimuli. Animals that did not learn the second discrimination in 30 days were arbitrarily assigned a criterion score of 30.

Results

The response measure was the number of days, excluding the criterion day, required to attain the experimental criterion. The mean training criterion scores for the 72 *S*s divided equally among Groups C, R, and NR were 5.96, 6.21, and 6.88, respectively. The training data were subjected to an analysis of variance which indicated that none of the variances were statistically significant.[4] Thus, the assumption that the test groups did not differ prior to the test discrimination is justified.

The test criterion scores were skewed, and the subgroup variances were extremely heterogeneous; therefore, the nonparametric *U* test (**7**) was used to assess the differences.[5]

The results indicate that Group NR shifted faster than Group R ($P < .05$), and that both of these groups were significantly more retarded than Group C ($P < .01$).

[4] Detailed descriptions of statistical treatments are presented elsewhere (**3**) and are available from the author.

[5] All the probabilities were doubled so that the results are consistent with a two-tailed hypothesis.

TABLE 1

ANALYSIS OF VARIANCE OF PERSISTENCE SCORES

Source	*df*	Mean Square	*F*
1. Treatments (R-NR)	1	70.08	10.90*
2. Conditions (BW-CN)	1	10.08	1.57
3. Interaction	1	0.08	<1.00
4. Within groups	44	6.43	

* $P < .01$.

The test discrimination was learned more readily when black-white was relevant than when chains-no chains was relevant ($P < .05$); however, the differences between Groups C, R, and NR were in the same direction under either stimulus condition.

A persistence score was computed for each *S* of Groups R and NR. This score was determined by the number of test days before the frequency of the correct *training response* fell below 80% on any one day.[6]

The analysis of variance presented in Table 1 indicates that the mean score of 4.58 for Group NR is reliably higher than the mean score of 2.16 for Group R. Although the *S*s of Group R abandoned the appropriate training responses more readily than Group NR ($P < .01$), they were slower to learn the test discrimination.

EXPERIMENT II

The plan of this experiment is quite similar to the previous one. Animals of the two experimental groups and the control group undergo the same training and testing procedures as their counterparts in the first experiment. There is, however, an important difference in the stimulus situation which prevails throughout the test trials. For all groups, the

[6] The same analysis was carried out on the number of days before the correct training response fell to 50%. The results were entirely consistent with those presented here.

irrelevant test cues are presented in successive configurations, i.e., one of the irrelevant cues appears on both sides on each test trial. This minimizes the possibility of differential responses on the basis of these cues; therefore, there should be no adventitious reinforcement for either experimental group.

The elimination of the partial reinforcement leads to a more critical test of the two theoretical formulations. The sequential S–R theory predicts that Group R should shift more rapidly than Group C which, in turn, should shift more rapidly than Group NR. The single-unit S–R theory can now be invoked without the necessity of making any assumptions about the effects of partial reinforcement. This theory predicts that Group NR and Group C should not differ significantly, but both should shift faster than Group R.

Method

Subjects.—The *Ss* were 79 experimentally naive, male, albino rats of the Wistar strain. Seven *Ss* were eliminated. Three could not learn the first discrimination, three would not perform, and one was lost through *E*'s error. At the beginning of the experiment *Ss* were 80–120 days of age.

Apparatus.—The discrimination apparatus has been described in Exp. I; however, there were two additional stimulus boxes. While metrically identical with the others, one had both black alleys and the other had both white alleys.

Procedure.—Exploration, pretraining, and training were identical with the analogous procedures in Exp. I.

The test situation depended upon the cue and the group to which the *S* was assigned in the first discrimination. For Group NR, the irrelevant training cues became relevant test cues; however, either the formerly positive or the formerly negative training cue was present in both stimulus alleys on any test trial. For example, if *S* had been shifted from chains-positive to black-positive, the stimulus alleys both contained chains or both contained no chains. For Group R, one of the two irrelevant training cues, which were also irrelevant test cues, was present in both stimulus alleys on any test trial. More specifically, if *S* had been shifted from chains-positive to no chains-positive, the stimulus alleys were both white or both black. The number of presentations under each condition was equalized. It is apparent that, during the test trials, the nonreversal *Ss* could no longer maintain their training response on the basis of the relevant training cues.

Results

The mean training criterion scores of the 72 *Ss* divided equally among Groups C, R, and NR were 6.46, 4.62, and 4.62, respectively. These training scores were subsequently transformed to meet the requisite assumptions for the analysis of variance. The analysis indicated that there are no significant differences among the test groups on the first discrimination.

The mean test criterion scores for Groups C, R, and NR were 4.58, 9.16, and 4.16, respectively. The test criterion scores were transformed to correct for heterogeneity of variance and correlation between means and variances. The analysis of the transformed scores is presented in Table 2. The results show that test learning is facilitated ($P < .05$) when the black-white cues are relevant. The treatments effect is significant at the .01 level, and *t* tests indicate that Group R is significantly retarded ($P < .01$) when compared with Groups NR or C. The difference between the latter groups is not significant.

TABLE 2

ANALYSIS OF VARIANCE OF TEST CRITERION SCORES TRANSFORMED TO LOG $(X + 1)$

Source	*df*	Mean Square	*F*
1. Treatments (C-R-NR)	2	0.881	8.89**
2. Conditions (BW-CN)	1	0.618	6.24*
3. Interaction	2	0.049	<1.00
4. Within groups	66	0.099	

* $P < .05$.
** $P < .01$.

Discussion

According to the sequential S-R formulation (**4, 5**) there are two stages of discrimination learning. In the first stage, mediating responses and cues are developed which make the relevant stimuli more discriminable. The second stage involves the association of instrumental behavior with these mediating processes. A reversal shift should be facilitated because the same cues are relevant in both discriminations; however, the nonreversal shift should be retarded because different cues are relevant in each discrimination. The results of both experiments indicate that the reversal shift leads to negative, rather than positive, transfer. Even more damaging to the sequential S-R formulation is the finding that the nonreversal group shifts more rapidly than the reversal group.

The application of the sequential S-R theory to discrimination learning in rats finds its main empirical support in the work of Lawrence (**4, 5**). In his first experiment (**4**), the results on the successive discrimination (first test) supported the sequential theory, but the results of the simultaneous discrimination (second test) showed a significant interaction which prohibited any general conclusions. Lawrence's second experiment (**5**) involved three stages. In the first stage, *Ss* learned a successive discrimination involving two cue dimensions (black-white and chains-no chains). One dimension was relevant and the other irrelevant. In the second stage, all *Ss* learned a simultaneous discrimination which could be mastered on the basis of either one or both sets of cues, i.e., both dimensions were relevant. In the third stage, the sequential hypothesis was tested in three ways: (*a*) the two cues which had been positive in the second stage were put in opposite alleys (opposition test); (*b*) half of the *Ss* in each group relearned the simultaneous discrimination with only one set of cues present (relearning test); and (*c*) the remaining *Ss* also had only one set of cues present but the cue values were reversed (reversal test). Lawrence interpreted his results as indicating that the cues which were relevant in the first stage were the most important determiners of the test responses. Even the reversal test was learned more rapidly when these cues were present, and this finding has been cited as being particularly damaging to a single-unit theory (**5; 8,** p. 453).

Inspection of Lawrence's data suggests an inconsistency in his interpretation. The *Ss* that were expected to approach chains or no chains on the opposition test responded at a chance level, and this showed up as a significant interaction chi square. The subgroup means for the relearning and reversal tests reveal that the results were *opposed* to the sequential theory when the test cues were chains-no chains.[7] It is possible that the results show no more than a preference for responding on the basis of black-white cues. Only the results of the reversal test on the black-white dimension are consistent with the sequential theory. Lawrence used the "correction-punishment" technique, and it is conceivable that a high response tendency to the incorrect test stimulus could facilitate reversal due to the rapid accumulation of inhibitory potential (cf. **1**). A more adequate assessment of this experiment could be made if a control group had been included. This group could have learned a neutral discrimination in the first stage and then proceeded to the same training and testing as the experimental groups.

The results of both experiments in the present study are consistent with the single-unit S-R theory. According to

[7] This would suggest the possibility of a significant interaction between experimental treatments and stimulus conditions on the relearning and reversal tests; however, the analysis of variance was carried out on transformed scores and showed no interaction. Lindquist (**6,** p. 216) notes, "Unless the main effect of one of the treatments is nil, the hypothesis of no interaction cannot be true for both the original and transformed measures." In this case there may be an interaction on Lawrence's original scale.

this formulation, the response is directly determined by the excitatory potentials of the various stimulus components. This theory accurately predicted the negative transfer for both types of shift in Exp. I, and the retardation of the reversal shift in Exp. II. For the explanation of discrimination learning in inarticulate organisms, it seems premature to abandon the single-unit S-R theory in favor the sequential S-R theory.

Summary

This study included two experiments which were designed to determine the effects of reversal and nonreversal shifts upon discrimination learning in white rats. In each experiment, two experimental groups (reversal and nonreversal) of 24 *Ss* were trained on a simultaneous discrimination which involved two sets of cues. A control group of 24 *Ss* was trained on a different discrimination involving one set of cues.

In Exp. I, the test discrimination involved the same sets of cues as the training discriminations. For Group R, the relevant training cues were still relevant, but the cue values were reversed. For Group NR, the irrelevant training cues became the relevant test cues. If Group NR persisted in their training responses, they received fortuitous partial reinforcement. The control *Ss* were tested on the same discriminations as the experimental *Ss*. In Exp. II, the partial reinforcement was eliminated by having one of the irrelevant test stimuli present in both alleys on any test trial. Except for this change in the test situations the two experiments were identical.

The results of Exp. I indicated that both types of shift result in negative transfer; however, the nonreversal shift was accomplished more readily than the reversal shift. In Exp. II, there was no difference between the test performances of the control and nonreversal groups, but the reversal group was significantly retarded. The results were interpreted as supporting the single-unit S-R theory of Spence, while being inconsistent with the sequential S-R theory of Lawrence.

References

1. Blum, R. A., & Blum, J. S. Factual issues in the "continuity" controversy. *Psychol. Rev.*, 1949, **56**, 33–50.
2. Ehrenfreund, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. physiol. Psychol.*, 1948, **41**, 408–422.
3. Kelleher, R. T. Discrimination learning as a function of reversal and nonreversal shifts. Unpublished doctor's dissertation, New York Univ., 1955.
4. Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
5. Lawrence, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
6. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
7. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
8. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univ. Press, 1953.
9. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
10. Spence, K. W. Continuous versus noncontinuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **47**, 271–288.
11. Wyckoff, L. B., Jr. The role of observing responses in discrimination behavior. *Psychol. Rev.*, 1952, **59**, 437–442.

(Received May 17, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

SOME SOURCES OF DIFFICULTY IN SOLVING SIMPLE PROBLEMS [1]

JACQUELINE J. GOODNOW[2] AND THOMAS F. PETTIGREW

Harvard University

One of the curious features of human discrimination learning and problem solving is the amazing length of time many *Ss* take to notice that a simple and regular pattern of events is occurring. Prentice's experiment (**16**) is a good example. His *Ss* chose between two designs on each trial. One design was always a square and the other a circle; the size of either could be large or small, the position on the left or the right, and the color of the figure or the background could be either black or white. The event pattern to be learned was that circle was the correct choice when the design on the right had a black background, and square the correct choice when the background was white. The *Ss* were allowed to take their time in making a choice, but still for 20 *Ss* the mean trials for learning was 41 (SD = 18). And 13 additional *Ss* had not learned at all by the end of 100 trials. A similar wide range of scores, with some extremely high scores, has been reported by Gellerman (**3**) with a double alternation temporal maze, and by Marquart (**13**) with a discrimination pattern comparable to Prentice's.

What lies behind this slowness in learning a relatively simple event pattern? In all the experimental situations cited, *S* is required to make a predictive response to each presentation of the stimulus. Our observations in a previous experiment (**5**) have led us to believe that the core difficulty for pattern learning is the presence of these responses. The *S*'s attention is directed more toward his own responses and hypotheses, and their successfulness, than toward the events (i.e., toward the alternative which turns out to be correct). This response-hypothesis orientation has a variety of effects. The *Ss* use their choices as direct tests of specific hypotheses rather than as tools for data gathering with hypothesis testing held in abeyance. As a result, the information gathered is related to a specific hypothesis and, if the latter should prove to be wrong, it is only with difficulty that the information can be transformed and made relevant to another hypothesis. As a rule, *Ss* do not transform information but start from scratch again with their next hypothesis. Furthermore, *Ss* frequently become so preoccupied with their level of success and failure that they can pay attention to little else. Finally, and most important, in the course of directly testing one hypothesis after another, the variability of their own responses obscures any event pattern. It is as if one tried to make observations in a scientific inquiry without a standard set of operations.

If this response-hypothesis orientation is a major difficulty in the learning of event patterns, then learning should be facilitated by any

[1] One of a series of studies comprising the Cognition Project, Laboratory of Social Relations, Harvard University. The authors wish to express their appreciation of advice given by Dr. Jerome S. Bruner and Dr. Ray Hyman, and assistance in running the experimental groups given by Dr. L. Minturn.

[2] Now at Walter Reed Army Institute of Research.

experience which leads *S* to pay attention to the events rather than to his own responses and especially by any experience which leads him to keep his own responses so consistent that change or regularity in the event pattern will stand out clearly. It is with this hypothesis that the present experiment is concerned.

This hypothesis does not rest entirely upon our own observations. Thorndike (**17**) and Adams (**1**) long ago noted the advantages of "attention, often correlated with a lack of vigor . . . [with] the absence of a fury of activity" for successful performance by cats in puzzle boxes (**17,** p. 46). Prentice comments on the "random and repetitive nature of solutions suggested by those who never succeeded in solving the problem" (**16,** p. 189). Heidbreder (**6**) has been most explicit about the predominance of participant behavior or direct hypothesis testing in problem solving, and together with Gellerman (**3**) and Brown and Buel (**2**) has noted the effectiveness but rare occurrence of *S*s using a series of identical responses as a deliberate tool in event observation.

Method

Apparatus.—The apparatus and general procedure have been described in detail elsewhere (**4**) and are the same as described in the previous experiment (**5**). Briefly, *S* plays a kind of two-armed bandit, with a choice on each trial between pressing a left key and a right key. The *S* activates the machine by inserting a chip, and then presses the key which he thinks will pay off. If *S* selects the correct key, the bandit drops a chip into a payoff slot. This is counted as a win. If no chip comes down, this is counted as a loss. The *S* is staked by *E* to 250 chips. At the end of the experiment, *S* is paid the difference between the number of times he has won and the number of times he has lost (one cent for each unit). The bandit is operated by *E* from an adjoining room. All *S*s are led to believe, however, that the machine is automatic. They are given only one piece of information about the way the bandit pays off. This is that one key or the other, never neither nor both, will pay off on each bet.

Design.—For all four groups of *S*s, the final pattern to be learned is LR (single alternation). For three groups, the order of experience is (*a*) learning an initial pattern to criterion (either LR, LL, or LLR); (*b*) 70 trials of "random" 50:50 payoff, and (*c*) learning the final pattern LR to criterion. For the fourth group (the control group), Step (*a*) is omitted.

The initial pattern is presented to *S* until he meets the criterion of 15 consecutive errorless trials, unless he fails to begin meeting this criterion within the limit of 100 trials. All *S*s move from one phase to the other without any time break. All *S*s are told that the experiment is a study of gambling and no information is given about possible patterning.

These three initial patterns (LR, LL, and LLR) were chosen as a check on the following possibilities: (*a*) Any prior experience of regularity will lead to an expectation of order or regularity and to closer attention to the environmental events. This possibility is an extension of Hilgard's arguments (**7**). If it holds, all groups experiencing initial regularity, whatever its form, should learn final LR faster than the control group. (*b*) Prior experience with one pattern leads to *S*'s being alerted for recurrence of the same pattern. This possibility would follow from Lawrence's study of acquired distinctiveness (**10**) and from our own previous experiment (**5**). In such a case, the group starting with LR should show faster final learning of LR. (*c*) Prior experience with a pattern leads to faster final learning when the prior experience establishes a consistent mode of response against which the final LR pattern stands out in relief. If this is so, the LL group should show faster final learning of LR than does the control.

The argument behind this third possibility rests on the special nature of the *arrangement of trials in the 50:50 period.* The event probabilities add up to 50:50 over the series of 70 trials but depart from it within each block of 10 trials. In the first block of 10 trials, the events (the alternatives turning out to be correct) are in the form 8L:2R, in the second block 4L:6R, and in subsequent blocks 6L:4R, 3L:7R, 7L:3R, 5L:5R, and 2L:8R. We used this arrangement because it gives longer runs of events than the usual random arrangement within restricted blocks, and, because of these longer runs, it is especially conducive to maintaining or disrupting consistent modes of response established by the initial patterning experience.

The group for which a long-run 50:50 arrangement maintains a consistent mode of

response already established is the LL group. We know from the previous experiment (**5**) that experience with LL followed by long-run 50:50 gives rise to a strategy of "win-stay, lose-shift" throughout the 50:50 phase. The *S*s stay with a winning key for as long as it pays off and shift from it only after it has ceased to pay off for one or two trials. It is this consistent mode of response which, we expect, will facilitate learning final LR by giving rise to such a consistent form of error that the event pattern will be easily grasped. The way in which this can occur is illustrated below.

Key Paying Off	*Response**	*Response***
L	L	L
R	L	L
L	R	L
R	L	L
L	R	L
R	L	L

* With strategy of win-stay; shift after losing once.
** With strategy of win-stay; shift after losing twice.

When *S* shifts after losing once, and the event pattern is LR, he is always one step behind the machine. When he shifts only after losing twice, he is led into a series of identical responses and wins on every second trial. Either way, the event alternation should be highly discriminable.

The LL group is the only group where we would expect the long-run 50:50 arrangement to sustain nicely the strategy established by the initial experience. Whatever the form of response to which experience with LR or LLR predisposes *S*, it is unlikely to be sustained by an arrangement where there are long runs of payoff on each of the two keys.

Subjects.—The *S*s were Harvard undergraduates, almost exclusively freshmen. Each group contained 20 *S*s. They received a base pay of 75 cents plus any winnings from the bandit.

Results

Speed of initial learning.—Table 1 shows the distribution of trials required to begin meeting the criterion of 15 consecutive errorless trials. The interesting feature is the increasing difficulty of the pattern learning, from LL to LR to LLR. One *S* (out of 20) fails to learn initial LR. Five *S*s (out of 20) fail to learn LLR. A wide spread of scores with some failures is in line with results reported by *E*s who used more complex event patterns (**13, 16**), but we had not expected the failure rate in learning LLR to be so high. The difficulty itself is intriguing, because *S*s consistently report that they are looking for a system based on such sequences as "two lefts, three rights," etc. Their hypotheses are as a rule far more complicated than the patterns to be learned; LLR is the closest in complexity to what they are reportedly searching for and is yet the slowest pattern to be learned; both LL and LR, especially LL, are greeted with a certain dubiety because they seem "so simple," and yet the least expected, LL, is the fastest learned. The source of difficulty does not appear to lie in the pattern meeting *S*s' expectations but rather in the ease with which an event pattern can be masked by response variability.

One point should be noted about the *S*s who are especially slow in

TABLE 1

Distribution of Trials to Criterion in Initial Pattern Learning

Group	*N*	Trials to Criterion and % of *S*s in Each Group						Median
		0–20	21–40	41–60	61–80	81–100	100 + (Failure)	
LL	20	100%	—	—	—	—	—	6.0
LR	20	50%	25%	10%	—	10%	5%	21.0
LLR	20	20%	25%	10%	10%	10%	25%	49.0

TABLE 2

MEDIAN TRIALS TO CRITERION OF 15 CONSECUTIVE ERRORLESS TRIALS IN LEARNING INITIAL AND FINAL LR

Group	Order of Events			Median Trials to Criterion		
	Initial	Interpolated	Final	Initial Pattern	Final Pattern	*N*—Final Pattern
1	LR	50:50	LR	21.0	8.0	19
2	LL	50:50	LR		9.5	20
3	LLR	50:50	LR		12.5	15
4	—	50:50	LR		17.0	20

learning. One might expect that these *S*s would be the ones who had given up and had ceased looking for an event pattern. Such *S*s are rare in our sample, although they do occur. The more general rule is that, like Adams' cats (**1**), it is the fastest learners who are "uncommonly sluggish." The slow learners are the *S*s who are the most active and participant. They form hypotheses rapidly, usually prematurely, and they test one hypothesis after another. They learn little from the implications of past experience and see only the relevance of a result to the hypothesis being tested at that time. Learning with these *S*s occurs only when they "run out of ideas." As one *S* put it: "I couldn't think of any more systems there could be, so I just fed some chips in without anything special in mind, just watching, and there it was." This is Heidbreder's (**6**) primary form of spectator behavior in its most beneficial guise.

The rate of failure in learning initial LLR calls for caution in interpreting the scores for final learning by members of that group. The slow learners may well have been screened out by the initial learning requirement.

Speed of final learning.—Median trials for final learning of LR are shown in Table 2 and the distributions of scores in Table 3. The result of most importance is the advantage for final LR learning given by initial experience with either LR or LL, and the lack of advantage given by initial experience with LLR.

The LL and LR groups enjoy an obvious superiority over the control group (no prior experience of patterning). Median trials to criterion are 8.0, 9.5, and 17.0, respectively. The Mosteller-Bush variation (**14**) of the Mann-Whitney test yields *z* scores of 3.53 in a comparison of the LR group with the control, and 2.07 in a comparison of the LL group with

TABLE 3

DISTRIBUTION OF SCORES FOR LEARNING LR

Condition	Trials to Criterion and % of *S*s in Each Group							
	0–10	11–20	21–30	31–40	41–50	51–90	91–100	100 +
Final LR (Gp. 1)	68%	32%	—	—	—	—	—	—
Final LR (Gp. 2)	55%	20%	15%	10%	—	—	—	—
Final LR (Gp. 3)	33%	40%	20%	7%	—	—	—	—
Final LR (Gp. 4)	25%	25%	20%	10%	10%	—	5%	5%
Initial LR (Gp. 1)	35%	15%	10%	15%	10%	—	10%	5%

the control. A z of 1.96 is regarded as grounds for rejecting the null hypothesis.

The medians for the LLR group and the control (12.5 and 17.0, respectively) suggest that some advantage is given by initial experience of LLR, but this is not supported by inspection of the distributions in Table 3 or by the z score of 1.25. It is possible that a much larger sample might show a significant advantage accruing from initial experience with LLR. It is far more likely that the suggestion of advantage after LLR is the result of the slower learners in the group having been screened out by failure to learn the initial pattern.

These results enable us to reject the first of the three hypotheses discussed in connection with the design. This is that experience with *any* form of prior patterning will facilitate later pattern learning because it sets up a general expectation of order or regularity. This hypothesis does not fit the lack of facilitation by initial LLR.

The advantage given by initial LR experience fits with the hypothesis that prior experience confers distinctiveness, in Lawrence's sense (**10**), upon a particular pattern, or that *Ss* expect a pattern previously experienced to recur in later experience with the same stimulus material. Distinctiveness or expectation of recurrence alone, however, will not cover the advantage given for final LR learning by initial LL experience. For this effect of prior experience, we want to demonstrate that the advantage arises from the consistent mode of response established by the initial LL pattern and sustained by long-run 50:50.

Sequential forms of response during the 50:50 phase.—As we expected, the only group showing a highly consistent mode of response during the 50:50 phase is the LL group, and its mode of response is such that final LR stands out clearly against it.

The method for determining response structuring is as follows. We distinguish among four kinds of preceding trial in terms of the choice made (left or right key) and its outcome of win or loss. We then determine for each group the number of times that each kind of trial is followed by a choice of left and the number of times it is followed by a choice of right. The ratio of left-to-right choices after each kind of trial is compared with the ratio of total left to total right choices. Any significant deviation from this standard is taken as evidence of a "strategy," i.e., a way of choosing between two alternatives in the light of the previous trial. The extent of the deviation is taken as evidence of the consistency with which the strategy is followed.

TABLE 4

INCIDENCE OF LEFT AND RIGHT CHOICES, AND PROPORTION OF LEFT CHOICES, FOLLOWING FOUR KINDS OF PRECEDING TRIALS
(Data for Trials 11–70 of the 50:50 Period)

Group	Event on Preceding Trial and Subsequent Choice															*N*
	Chose L and Won			Chose L and Lost			Chose R and Won			Chose R and Lost			Total Choices			
	L	R	%L	L	R	%L	L	R	%L	L	R	%L	L	R	%L	
1 (LR)	126	122	50%	175	153	53%	126	153	45%	153	132	54%	580	560	51%	19
2 (LL)	249	51**	83%	201	164	55%	71	203**	26%	134	127	51%	655	545	55%	20
3 (LLR)	133	92	59%	164	84*	66%	114	117	49%	62	134*	32%	473	427	53%	15
4	206	91*	69%	194	156	55%	113	177*	39%	133	130	50%	646	554	54%	20

* Significant at 1% level.
** Significant at .1% level. Significance levels established with binomial probability paper (**15**).

The results of this analysis are presented in Table 4. The only group showing extensive use of a strategy is the LL group. The strategy used is one of repeating a successful choice and the extensiveness of its use is shown by the ratio of 5 to 1 in favor of repeating a successful left choice and 3 to 1 in favor of repeating a successful right choice. The control group (no initial pattern) shows a significant but less extensive tendency to use the same strategy of repeating a successful choice. This is a group which is apparently led by the experience of long runs in the 50:50 period to repeat successful choices. As in the previous experiment (**5**), however, these *S*s do not persist with a successful choice until the run is finished. Instead they attempt to switch to the opposite key before they have lost and at a point where they estimate that the run is about to finish. This same tendency to shift keys before an event change occurs is found in the LLR group, where *S*s develop a strategy of repeating a previously unsuccessful choice, but not persisting with it to the point where the key starts to pay off.

As we expected, the strategy of regularly persisting with a previously successful choice provides a consistent form of response against which the final pattern of LRLR stands out clearly. Persistence with a previously successful choice carries with it some provision for what to do when the run is finished and the choice leads to a loss. Some subjects adopt as their usual policy one of switching immediately after one unsuccessful choice. When these *S*s meet the pattern of LRLR with this policy, the result is a consistent pattern of error. In the words of one *S*: "I found I was always one step behind the machine, so I knew it must be alternating." Other *S*s adopt as their usual policy one of persisting with a previously successful choice until that key has lost twice, and then switching to the opposite key. When the pattern of LRLR is met with this policy, the result is a run of left or right choices with every second choice successful. As one *S* remarked: "I was staying with left until it lost twice in a row, and I found I was winning every second time, so the machine had to be alternating."

We have stressed the presence of a consistent form of response as a way of explaining the speed of final learning in the LL group and the absence of a consistent form of response as a way of explaining the slowness of final learning in the LLR group. This interpretation is strengthened by the results obtained with a group not so far reported. This group started with LR and ended with LL. Here again we found no consistent mode of response during the 50:50 period and no facilitation of the final learning.

It should be noted that we do not find within groups any relationship between degree of response structuring and speed of final learning. This lack of correlation is the result of three factors. One is that, beyond a certain point, increases in response consistency do not seem to give rise to any further advantages for final learning. The second is that measures of response consistency, in order to be based on a sufficient sample of responses, have to be taken over almost the entire series of the 50:50 phase, but response consistency, in order to be effective, may need only to occur toward the end of the 50:50 phase and in meeting the final LR pattern. The third factor is that occasionally response consistency rep-

resents a withdrawal from the problem rather than a tool for event observation; *S* has abandoned belief in the possibility of finding a regular event pattern or of improving his performance, and settles for a form of response which enables him "to do well enough." This source of response consistency is rare among our *S*s, but its occasional occurrence, together with the other two factors limits the possibility of a correlation within groups between degrees of response consistency and speed of final learning.[3]

Discussion

We are led to suggest two extensions to the concept of discriminability proposed by Longenecker, Krauskopf, and Bitterman (**11**) to cover the differences they found in speed of response to changes in event patterning. These extensions take the form of suggesting two ways by which the discriminability of event patterns can be enhanced or obscured.

When the initial and final patterns are the same, the initial experience may lend a useful distinctiveness, in Lawrence's sense (**10**), to the initial pattern—a distinctiveness buttressed by the fact that *S*s do expect patterns previously found important to recur in later experience with the same stimulus material. It may even be that prior experience confers distinctiveness not only upon one particular pattern but also upon similar patterns, and that the extent to which later learning is facilitated will be a function of the degree of similarity between initial and final patterns. This latter relationship is not likely to be a simple one, however, since *S*s expect change to proceed not only along a dimension of similarity (e.g., all payoffs on one side or the other, or all in some form of alternation), but also in the direction of increasing rather than decreasing complexity.

It is when the initial and final patterns are not the same that the second source of discriminability becomes important. This source is the consistency of the mode of response established by prior experience. Response consistency provides a standard set of operations for *S* to follow and a steady background against which the event pattern can stand out in relief.

These are certainly not the only two sources by which an event pattern can be made more discriminable. Any learning set or any experience which leads *S* to become event oriented rather than response-hypothesis oriented should serve the same function. Instructions such as "don't be blind" (**12**), training in suspending judgment until all the data are in (**18**), experience which lowers one's level of anxiety and concern with questions of success and failure (**9**)—all these have the function of directing *S*'s attention toward the events rather than toward his own responses and hypotheses. We have chosen to emphasize the role of response consistency partly because of its intrinsic significance and partly as a counterweight to the tendency to regard response consistency solely as stereotypy and therefore somehow disfunctional (**13**). Keeping one's own responses constant is as important in experimental pattern learning as using standard measures is in scientific problem solving.

Summary

Each of three groups of *S*s starts by learning that there is a consistent pattern of payoff on a two-choice gambling device—either LR, LL (all left), or LLR. Each *S* is then shifted to a random schedule of payoff for 70 trials, and finally a pattern of LR is introduced for all groups. A control group experiences no initial pattern but starts with the random period. Experience with initial LR or with LL, but not with LLR, leads to faster final learning of LR. The results are explained in terms of previous experience making the final pattern more discriminable, with discriminability strongly in-

[3] To test for such correlation, we used Hyman's (**8**) procedures, allowing measures of extent of contingency between response on one trial and (*a*) the response on the previous trial, (*b*) the alternative correct on the previous trial, and (*c*) the combination of response and outcome on the previous trial.

fluenced by *Ss* learning to adopt a consistent mode of response against which an event pattern stands out clearly rather than being masked by response variability.

References

1. Adams, D. K. Experimental studies of adaptive behavior in cats. *Comp. psychol. Monogr.*, 1929, **6**, No. 1 (Whole No. 27).
2. Brown, W., & Buel, J. Response tendencies and maze patterns as determiners of choice in a maze. *J. comp. psychol.*, 1940, **29**, 337–399.
3. Gellerman, L. W. The double alternation problem: II. The behavior of children and human adults in a double alternation temporal maze. *J. genet. Psychol.*, 1931, **39**, 359–392.
4. Goodnow, J. J. Determinants of choice in two-choice situations. *Amer. J. Psychol.*, 1955, **69**, 106–116.
5. Goodnow, J. J., & Pettigrew, T. F. Effect of prior patterns of experience upon strategies and learning sets. *J. exp. Psychol.*, 1955, **49**, 381–389.
6. Heidbreder, E. An experimental study of thinking. *Arch. Psychol.*, 1924, **11**, No. 5 (Whole No. 73).
7. Hilgard, E. R. The role of learning in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception—an approach to personality*. New York: Ronald, 1951. Pp. 95–120.
8. Hyman, R., Potter, M. C., & Dinwoodie, W. D. Personality correlates of behavioral stereotypy. *J. abnorm. soc. Psychol.*, in press.
9. Keister, M. E. The behavior of young children in failure. In R. G. Barker, J. S. Kounin, & H. F. Wright (Eds.) *Child behavior and development*. New York: McGraw-Hill, 1943. Pp. 429–440.
10. Lawrence, D. H. Acquired distinctiveness of cues. I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
11. Longenecker, B. D., Krauskopf, J., & Bitterman, M. E. Extinction following alternating and random partial reinforcement. *Amer. J. Psychol.*, 1952, **65**, 580–587.
12. Luchins, A. S. Mechanization in problem solving: the effect of *Einstellung*. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).
13. Marquart, D. I. The pattern of punishment and its relation to abnormal fixation in adult human subjects. *J. genet. Psychol.*, 1948, **39**, 107–144.
14. Mosteller, F., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.) *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 289–334.
15. Mosteller, F., & Tukey, J. W. The uses and usefulness of binomial probability paper. *J. Amer. Stat. Ass.*, 1949, **44**, 174–212.
16. Prentice, W. C. H. Continuity in human learning. *J. exp. Psychol.*, 1949, **39**, 187–194.
17. Thorndike, E. L. *Animal intelligence*. New York: Macmillan, 1911.
18. Thorndike, E. L. How children learn the principles and techniques of problem-solving. Chicago: *Natl. Soc. Stud. Educ., 49th Yearb.*, Part I. Pp. 196–216.

(Received June 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

EFFECTS OF STIMULUS SIMILARITY ON DISCRIMINATION LEARNING

CHARLES C. SPIKER[1]

Iowa Child Welfare Research Station

Increase in the difficulty of a discrimination task, produced by increase in the similarity of the discriminal stimuli, has been explained in terms of increased generalization of incorrect approach and inhibitory tendencies (**1, 3, 4**). Observations of children in discrimination problems have led this investigator to consider the possibility that increase in stimulus similarity increases task difficulty through the introduction of factors in addition to stimulus generalization. The present experiment is designed to test for the presence of such additional factors and to attempt to provide a measure of stimulus generalization that is not contaminated by them.

A successive discrimination task in which one or more of the available responses is always incorrect offers a means of isolating task difficulty contributed by stimulus generalization and that contributed by other factors. Increase in stimulus generalization would be expected to increase generalized (intralist) errors, but it would not be expected to increase the frequency of occurrence of responses that are never correct (extralist errors). On the other hand, if increase in stimulus similarity increases the potency of inattention, disinterest, frustration due to failure, etc., one would expect an increase in the frequency of both intralist and extralist errors.

[1] The writer is indebted to Dr. Irma R. Gerjuoy who participated in the preliminary experiments upon which this experiment is based. He also wishes to thank his research assistant, Miss Ruth B. Holton, for her aid in collecting the data.

METHOD

Apparatus.—The apparatus was a black box, 18 × 24 × 16 in. Six inches above a 3-button response panel was a 3-in. diameter aperture of flashed opal glass. White geometric figures (square, circle, or triangle) were painted around each of the buttons, each of which was approximately 7 in. from the other two. Two 75-w. projection lamps were mounted 6 in. behind the aperture, each enclosed in a light-proof reflecting container on the front of which a 2 × 2-in. glass color filter could be inserted. The *E* could turn on either one of the lamps and predetermine which of the three buttons would break its circuit. For each *S*, one button did not turn out either lamp. If *S* pushed the correct button, the lamp was turned out and a marble automatically delivered to *S*. Incorrect responses resulted in the disconnection of all buttons, thereby preventing more than one response per trial.

Experimental design.—Three experimental groups differed only in the discriminal stimuli presented. The stimuli were obtained by diffusing light through glass color filters (Corning Glass Works) onto the aperture. Group HS was presented with a yellow-green (No. 4015) and a green (No. 4308 superimposed on No. 3484). Group S was presented with the yellow-green and a blue-green (No. 3385 superimposed on No. 4303). Group D received the yellow-green and a deep blue (No. 5330). The stimuli in each pair are readily discriminated by adults with normal color vision. Within these groups, the position of the disconnected button, which was never correct, was controlled by randomly assigning *S*s to three subgroups.

Subjects.—Sixty-five preschool *S*s from the Iowa Child Welfare Research Station laboratories were randomly assigned to the experimental groups—25 to Group HS, 22 to Group S, and 18 to Group D. The *S*s ranged in age from 4 yr., 2 mo. to 5 yr., 8 mo. Some *S*s had previous experience with simultaneous and successive discrimination learning problems, but none had previous experience with the present apparatus.

Procedure.—A familiar *E* brought *S* to the experimental room and allowed him to select one of a number of dime-store toys which he would attempt to win. The *E* then demonstrated the two stimulus lights one at a time, instructing *S*

that marbles could be obtained by turning out the lights, and that the marbles could be used to buy the selected toy. The appropriate stimuli were presented in a prearranged order that was the same for all *Ss*. Each light appeared three times in successive blocks of six presentations, long series of alternations were avoided, and neither light appeared more than three times in succession. Button (position) preferences were verbally discouraged by *E*. All *Ss* received at least 30 stimulus presentations (trials); *Ss* who failed to attain 9 correct choices in 10 consecutive trials by Trial 30 were given additional trials until this criterion was attained or until 96 trials had been given. Eleven of the 65 *Ss* failed to reach this criterion in 96 trials.

Results

Analysis of the total error scores for the subgroups, which differed only in terms of the position of the disconnected button, revealed no consistent or significant differences among them. Accordingly, the data for these subgroups have been combined within each experimental group. The scores used in subsequent analyses are the number of errors involving the button that was connected (intralist errors) and the number of errors involving the button that was disconnected (extralist errors).

Table 1 shows the means of the two types of error for each of the three experimental groups. From inspection, it is apparent that both error types tend to increase in frequency with increase in stimulus similarity, and that the magnitude of the difference between the means of the two error types also increases with increasing stimulus similarity. For each group, the mean frequency of intralist errors is greater than that of extralist errors.

TABLE 1

Means of the Two Types of Error

Groups	*N*	Intralist Errors	Extralist Errors	Total Errors
HS	25	21.48	7.44	28.92
S	22	17.64	6.95	24.59
D	18	6.44	3.28	9.72
Combined	65	16.02	6.12	

TABLE 2

Analysis of Variance of Number of Errors to Criterion

Source	*df*	Mean Square	*F*	*P*
Ss:	64			
Stimuli (S)	2	1014.28	6.09	.005
Error (b)	62	166.47		
Within-*Ss*:	65			
Error Type (E)	1	3180.38	32.34	.001
E × S	2	314.50	3.20	.05
Error (w)	62	98.34		
Total	129			

The statistical significance of these trends was tested by a Lindquist Type I analysis (**2**) and by further tests of the simple effects. The results of the Type I analysis are presented in Table 2. Here it is apparent that the experimental groups differ significantly in terms of mean total errors, and that frequency of intralist errors is significantly greater than the frequency of extralist errors for the combined experimental groups. Furthermore, there is an interaction ($P < .05$) between the stimulus groups and the type of error. The interaction indicates that, with increasing stimulus similarity, there is an increase in the magnitude of the difference between the mean frequencies of the two types of error.

The data were analyzed as two simple randomized designs to determine the stimulus effects on the two types of error separately. For the intralist errors, the stimulus effects were significant at the 1% level ($F = 5.54$; $df = 2, 62$). For extralist errors, the stimulus effects were significant at the 2.5% level ($F = 4.25$; $df = 2, 62$). Using the t test, comparisons were made of pairs of groups.

For the intralist errors, the mean of Group D differs significantly from that of Group HS ($P < .01$) and from that of Group S ($P < .02$). The means for Groups HS and S do not differ significantly. For the extralist errors, also, the Group D mean is significantly lower than the Group HS mean ($P < .01$) and the Group S mean ($P < .02$), but the means for Groups HS and S are not significantly different. Clearly, an increase in stimulus similarity resulted in an increase of errors of both types.

Finally, t tests were made of the significance of the differences between means for intralist and extralist errors for the three experimental groups separately. Significantly more intralist errors than extralist errors were made by Groups HS and S ($P < .001$), but the difference for Group D was not statistically significant ($P > .05$).

Discussion

The greater frequency of intralist than of extralist errors presumably reflects the operation of stimulus generalization, which, in the present experiment, is expected to increase intralist errors without directly increasing the number of extralist errors. The interaction between stimulus similarity and type of error is interpreted as the result of an increase in stimulus generalization with increasing stimulus similarity. This interaction is somewhat more statistically reliable ($P < .01$) when only the extreme groups are considered. It would seem that the interaction provides an index of the amount of generalization that is not affected by other factors that may be associated with increased stimulus similarity.

The increase in frequency of extralist errors with increasing stimulus similarity is only indirectly due to increased stimulus generalization. A task made difficult by high generalization is made even more difficult through loss of interest due to increased number of failures; elicitation of inappropriate responses (orienting away from stimuli, asking to leave the situation, hitting buttons indiscriminately, etc.); and the development of fatigue due to longer experimental sessions. Qualitative observations suggest that such factors were more frequent among *S*s given highly similar stimuli than among those given distinctive stimuli.

The evidence indicates that errors on a discrimination learning task may provide an overestimate of the direct effects of stimulus generalization. Children's performance, at least, appears to be affected by stimulus similarity in a manner that can only partly be attributed directly to stimulus generalization.

Summary

Three groups of preschool children were given a successive discrimination learning task in which generalized (intralist) or nongeneralized (extralist) errors could be made. The conditions for the three groups differed only in the similarity of the discriminal stimuli. The results indicated that increased stimulus similarity resulted in greater numbers of both types of error and an increased difference between the means of the two types of error. These results were interpreted as reflecting increased stimulus generalization as a function of increased stimulus similarity. The increase in extralist errors with increased stimulus similarity was interpreted as reflecting the operation of factors in addition to stimulus generalization, e.g., decrease in interest, motivation, or attention, and increase in fatigue.

References

1. Hull, C. L. *A behavior system.* New Haven: Yale Univ. Press, 1952.
2. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
3. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
4. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

(Received June 15, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

CONSTANT-SUM JUDGMENTS OF FACIAL EXPRESSIONS [1]

TRYGG ENGEN AND NISSIM LEVY
Brown University

The main purpose of this study was to determine the potentiality of the constant-sum method in psychometric scaling. An answer was sought to the question as to whether or not the constant-sum method could be successfully applied to scaling of pictures of facial expression. The method deserves attention because it may yield ratio scales, but it has thus far been used only in traditional psychophysical experiments, i.e., judgments of lines and weights have been compared to physical scales (**1, 2, 3, 4**).

A second aim was to compare scales of facial expressions developed by the method of rating and the constant-sum method. For this purpose, there are available the ratings provided by Schlosberg (**8, 9, 10**) on the dimensions of pleasantness-unpleasantness (P–U), attention-rejection (A–R), and sleep-tension (S–T).

Method

Three group experiments were undertaken, and each employed the same experimental procedure. Fifteen Frois-Wittmann pictures were selected to sample eight values on each of Schlosberg's (**10**) three dimensions. Pictures 58, 41, 2, 26, 43, 51, 31, and 70 from the Frois-Wittmann series (**5**) were chosen to represent the P–U dimension, Pictures 41, 21, 51, 12, 43, 69, 26, and 28 the A–R dimension, and Pictures 52, 51, 67, 12, 70, 58, 36, and 34 for the S–T dimension. Some of the pictures represent more than one dimension, i.e., Pictures 41, 26, 43, and 51 occur in the samples of both P–U and A–R, and Pictures 58, 51, and 70 represent both P–U and S–T, and Picture 51 appears in all three dimensions.

Photographic slides were made of the 28 possible pairs of the 8 pictures representing each dimension, and these were projected on a screen for group presentation. The slides were presented in a random order with the one stipulation that the same picture could not appear in two successive slides. There were two presentations of each series of 28 slides, and these were counterbalanced for series and right-left position effects.

The same students in an introductory class in psychology served as *Ss* for the three experiments, which were one week apart. The *N*, in order of presentation of the dimensions, was 42, 37, and 39 for the P–U, A–R, and S–T dimensions, respectively. The *Ss* recorded their responses on prepared data sheets. They were given approximately 30 sec. for each judgment, which was ample time as determined by preliminary trials on other *Ss*. It should be noted that *Ss* complained about the difficulty of the task, although each dimension was carefully defined. The instructions emphasized that 100 points were to be divided between the members of each pair of pictures in terms of the relative amount of pleasantness (or attention or tension) expressed in the faces. The method was demonstrated with vertical lines drawn on the blackboard and with pictures from the Ruckmick (**7**) series representing two relative extremes of each dimension. The *Ss* were discouraged from using the 100-0 division for any of the pairs on the basis that these pictures probably did not represent the most extreme emotional expressions they could imagine. They were encouraged to use any other division of 100 they thought appropriate and not to restrict themselves to five or ten point divisions.

Results and Discussion

Scale values were developed from the *Ss*' point assignments by formulae presented by Comrey (**2**), and the results of these computations are presented in Table 1 along with the Schlosberg rating values for the same pictures. As far as ordinal characteristics are concerned, the results of the present three experiments are in almost complete agreement with those of Schlosberg. For example, it should

[1] This paper was presented at the 1955 meetings of the Eastern Psychological Association in Philadelphia.

TABLE 1

CONSTANT-SUM METHOD (CSM) AND RATING METHOD (RM) SCALE VALUES ON DIMENSIONS OF P–U, A–R, AND S–T

Picture numbers are Frois-Wittmann's (5).

P–U			A–R			S–T		
CSM	Picture Number	RM	CSM	Picture Number	RM	CSM	Picture Number	RM
1.0	58	2.0	1.0	41	5.0	1.0	52	2.6
1.1	41	2.1	1.1	21	4.1	1.1	51	2.2
1.4	2	2.7	1.3	51	3.2	2.6	67	2.6
1.7	26	5.0	1.6	12	4.7	3.2	12	5.0
1.7	51	5.0	1.8	43	5.0	4.1	70	5.1
1.9	43	5.0	2.1	69	5.2	6.7	58	6.0
2.9	31	6.2	4.2	26	7.0	7.2	36	8.4
5.6	70	8.4	6.2	28	7.6	7.9	34	8.6

Note.—The S–T data, which have not been published previously, were kindly supplied by Prof. H. Schlosberg. They had been obtained by the same rating method used for the other two dimensions with an N of 20.

be noted that some of the pictures (i.e., 26, 43, 51, 58, and 70), which had markedly different scale values on Schlosberg's three dimensions, likewise obtained different scale values on the three constant-sum scales. The tau rank-order correlations between the scales developed with the two methods are .96 for P–U, .92 for A–R, and .92 for S–T.

Only one large inconsistency was observed. This was Picture 41 on the A–R scale. However, the constant-sum method scale value is consistent with the Woodworth scale position given by Schlosberg (8) in 1941. This picture may well have been misplaced in Schlosberg's (9) 1952 study either through error or through some obscure series effects.

The present results demonstrate S's ability to describe facial expressions with numbers. The direct estimates of the ratio between adjacent stimulus pairs (e.g., A/B) are in good agreement with the indirect estimates (e.g., A/B = A/C ÷ B/C, etc.). The means of the differences between the six indirect and the one direct estimates were .05, .15, and .09 for slightly positively skewed distributions with SD's of .21, .33, and .25 for A–R, S–T, and P–U, respectively. In the case of the most variable dimension, this means that if the direct estimate of A/B was 2/1, which is reflected in a 67–33 point assignment, about 68% of the indirect estimates were between 64–36 and 70–30. The small variability is surprising in view of Ss' complaints about the difficulty of the task. This internal consistency compares well with Comrey's (2) results. Our reanalysis of his data obtained with the constant-sum method on judgments of lines yields a mean of .05 and an SD of .15, i.e., with a direct estimate of 67–33, 68% of the indirect point assignments were between 65–35 and 68–32.

The general good agreement between the scales developed by the two methods and the internal consistency of the constant-sum scales render it feasible to answer affirmatively the question which led to these experiments. That is to say, it has been demonstrated that the constant-sum method can be applied with success to psychometric dimensions. Finally, the high agreement between the constant-sum judgments and the

ratings lends evidence to the stability of Schlosberg's dimensions.

Summary

Fifteen of the Frois-Wittmann pictures of facial expressions were scaled by the constant-sum method and compared with Schlosberg's rating scales of the same pictures on his dimensions of pleasantness-unpleasantness, attention-rejection, and sleep-tension. The constant-sum scales were in good agreement with the rating scales as far as ordinal characteristics were concerned. Further, the constant-sum scales showed high internal consistency in terms of interratio agreements, although not quite as high as has been found with this method in psychophysical scaling of lines. The results provide evidence that the constant-sum method, which may yield ratio scales, can be extended to material more complex than lines and weights.

References

1. Baker, K. E., & Dudek, F. J. Weight scales from ratio judgments. *J. exp. Psychol.*, 1955, **50**, 293–308.
2. Comrey, A. L. A proposed method for absolute ratio scaling. *Psychometrika*, 1950, **15**, 317–325.
3. Engen, T. An evaluation of a method for developing ratio scales. *Amer. J. Psychol.*, in press.
4. Guilford, J. P., & Dingman, H. F. A validation study of ratio-judgment methods. *Amer. J. Psychol.*, 1954, **67**, 395–410.
5. Hulin, W. S., & Katz, D. The Frois Wittmann pictures of facial expression. *J. exp. Psychol.*, 1935, **18**, 482–498.
6. Metfessel, M. A proposal for quantitative reporting of comparative judgments. *J. Psychol.*, 1947, **24**, 229–235.
7. Ruckmick, C. A. A preliminary study of the emotions. *Psychol. Monogr.*, 1921, **30**, No. 3 (Whole No. 136). Pp. 30–35.
8. Schlosberg, H. A scale for judgment of facial expressions. *J. exp. Psychol.*, 1941, **29**, 497–510.
9. Schlosberg, H. The description of facial expressions in terms of two dimensions. *J. exp. Psychol.*, 1952, **44**, 229–237.
10. Schlosberg, H. Three dimensions of emotion. *Psychol., Rev.*, 1954, **61**, 81–88.

(Received June 27, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

PERCEPTUAL BEHAVIOR AS RELATED TO FACTORS OF ASSOCIATIVE AND DRIVE STRENGTH

JEAN ENGLER[1] AND JAMES T. FREEMAN

Iowa State College

Much of current behavior theory regards behavior as being determined by at least two general classes of variables. These are usually considered as associative or learning variables and drive variables. The most explicit statement of the functional properties of these variables has been expressed by Hull (**6**) who considers performance or response (*R*) to be a function of the interaction of habit strength (*H*) and drive (*D*). A number of studies have been carried out within the above framework in which anxiety has been manipulated as a drive variable.

Anxious *Ss* have been reported to condition more readily and to extinguish less readily on both the eyelid reflex (**14, 15, 16**) and the galvanic skin reflex (**4, 20**) than nonanxious *Ss*.

Farber and Spence (**2**) reported that the maze performance of anxious *Ss* was significantly poorer than that of nonanxious *Ss* with the greatest difference at the most difficult points. Taylor and Spence (**18**) noted that the performance of anxious *Ss* was inferior to that of nonanxious *Ss* in a learning situation involving the presence of competing response tendencies with the greatest difference for the more difficult task.

Montague (**8**) reported that anxiety increases the difference between stronger and weaker response tendencies in a learning situation, aiding performance if correct tendencies are dominant. Ramond (**12**) reported that in two-choice response type of paired-associate learning nonanxious *Ss* responded correctly significantly more often than did anxious *Ss* on those presentations for which the weaker of two possible responses was correct. There was no significant difference between performance of the anxious and nonanxious *Ss* on those presentations for which the stronger response was correct.

Recent trends in perceptual theory have indicated that much of perceptual behavior may be approached from the standpoint of learned response dispositions whose properties may be described in terms of general principles of associative learning. It should be noted that such formulations do not necessarily preclude the existence of motivational determinants. There would thus seem to be an apparent parallel between, say, Hullian behavior formulations on the one hand and those recently expressed by Postman (**9**), for example.

In a perceptual situation, the strength of response dispositions may be varied in two ways (among others): (*a*) by varying the frequency characteristics of the stimulus materials (**5, 7, 13, 21**), and (*b*) by manipulating the set of the *Ss* (**3, 10, 11**). The above studies, taken together, indicate that the higher the frequency of the stimulus words and the more restrictive the set of *S*, the easier will be the perceptual task (i.e., the greater the response disposition, the less the stimulus information needed, and hence the lower the recognition threshold). Past studies also indicate that in learning situations high-anxiety *Ss* are superior to low-anxiety *Ss* in performance on simple tasks while low-anxiety *Ss* are superior to high-anxiety *Ss* in performance on more complex tasks.

The present research is concerned with investigating perceptual recognition thresholds as a function of associative strength (or strength of the response disposition) as determined by word frequency and set, and as a function of drive (moti-

[1] Now at Northwestern University.

vation) as determined by anxiety. In line with the foregoing considerations, it is hypothesized that high-anxiety *Ss* will have lower recognition thresholds than low-anxiety *Ss* to high-frequency words under the set condition, i.e., these *Ss* should perform better in a simpler perceptual task. It is further hypothesized that the low-anxiety *Ss* will have lower recognition thresholds than high-anxiety *Ss* to low-frequency words under the no-set condition, i.e., these *Ss* should perform better in a more complex perceptual task.

Method

Subjects.—The *Ss* were 80 undergraduate students in introductory psychology classes at Iowa State College.

Apparatus.—The apparatus was a Harvard-type tachistoscope calibrated in .01 sec. The size of the exposure field was 10 × 10 in. with the stimulus words (typed in elite capitals) exposed through an aperture 1 × 3 in. centered in the exposure field. The intensity of illumination of the exposure field measured at the aperture was 6.0 ft.-candles. The distance of the exposure aperture from *S*'s eye was 28 in.

The size of the pre-exposure field was 10 × 10 in. with the fixation point (a cross ½ × ½ in.) centered in the pre-exposure field. The intensity of illumination of the pre-exposure field at the fixation point was 5.0 ft.-candles. The distance of the pre-exposure fixation point from *S*'s eye was 22 in.

All illumination intensity measurements were obtained directly from a Weston Illumination Meter, Model 603.

Experimental design.—Matched groups consisting of five high-anxiety and five low-anxiety females and five high-anxiety and five low-anxiety males served in each of four conditions. The conditions involved combinations of high- and low-frequency stimuli and set and no-set instructions.

Procedure for selecting Ss.—The MMPI A scale as developed and revised by Taylor (17) was administered to 200 male and 200 female students. The distributions of scores for males, for females, and for the total sample were plotted and compared to the distribution obtained by Taylor. All three distributions agreed in shape with Taylor's distribution, being positively skewed. The total-sample distribution had a significantly higher median than that reported by Taylor ($P = .03$) and the median and both quartiles were significantly higher for males than for females ($P = .02$).

The 20 males and 20 females with the highest A scores (24 to 40) and the 20 males and 20 females with the lowest A scores (2 to 10) were used as *Ss*.

Experimental procedure.—The same three practice words were used for each *S* and the mean recognition threshold for the second two of these was used as a practice measure for matching the four groups, which were then randomly assigned to the experimental conditions.[3] Following the three practice words *Ss* were given the experimental instructions and 10 stimulus words for the conditions to which they had been assigned. The words were presented in random order for each *S*. Recognition thresholds and pre-recognition hypotheses were recorded. Exposure proceeded by .01 sec. from zero to .20 sec., by .02 sec. from .20 to .30 sec., and by .03 sec. beyond .30 sec. The point of origin was varied throughout successive presentations in an attempt to conteract any response sets built up by *S*. Two consecutive correct responses constituted a recognition threshold which was recorded as the exposure time on the first of these responses. The reading was then converted to milliseconds for purposes of statistical treatment.

The instructions read to *Ss* were as follows: (*a*) General instructions (read for practice and no-set conditions): "I am going to show you some words one at a time and I want you to tell me everything you see or think that you see as quickly as you can. I will show the words to you for only a brief moment but I want you to tell me everything that you see or think that you see. The words will appear where the *X* is now. When I call ready, please look at the spot where the words will appear. The same word will be shown until you have correctly identified it twice in a row. Any questions? Ready." (*b*) Set instructions: "Some, but not all, of the words I'm going to show you now will be animal words, that is, words standing for animals. Your task is to recognize each word just as quickly as you can and report it to me. Remember, some but not all of the words will be animal words. Each word will be shown to you until you correctly identify it twice in a row. Any questions? When I call ready, please look at the spot where the words will appear. Ready."

Stimulus lists were constructed to be of homogeneous intralist frequency, that is, words within a given list were of essentially the same

[3] Five *Ss* were discarded in the process of matching the groups for sex and for high and low anxiety.

Thorndike-Lorge general word count (**19**). Each list consisted of four critical words (words consistent with set) and six neutral words (words not consistent with set). Length was held constant insofar as was possible. The practice words (medium frequency) were *brick*, *straw*, and *lunch*. The stimulus lists were as follows:

High-frequency		*Low-frequency*	
critical	*neutral*	*critical*	*neutral*
chicken	among	gazelle	align
horse	certain	opossum	corrode
rabbit	refuse	raccoon	crimp
sheep	remove	sloth	fluster
	there		humid
	under		terse

Results

For each *S* the mean recognition threshold for the four critical words and the six neutral words served as the basic data upon which an analysis of variance was performed. This analysis appears in Table 1. It may be seen that such variables as frequency, words, and words by set were all highly significant, which is consistent with a number of previous studies. Of more critical importance to the tests of the hypothesis of the present study, however, is the analysis of the interactions, particularly those involving set, anxiety, and frequency. A graphical representation of the Set × Frequency interaction, which approached statistical significance (*F* of 3.60; 3.98 required at .05 level), is presented in Fig. 1. The general effects of frequency (high vs. low) upon over-all thresholds are evident, and it is to be noted that the effect of set upon the critical words is more apparent in the case of the low-frequency material ($t = 3.07$, $P = .01$). For high-frequency material, the effect of set upon critical words was not significant although the effect of set was associated with significantly higher thresholds ($t = 4.56$, $P < .01$) for the neutral words. In short, it would seem that the effect of set in the direction of lower thresholds was much more pronounced for low-frequency material while making little if any difference in the case of high-frequency material, which corroborates a finding previously noted by the writers (**3**).

TABLE 1

Analysis of Variance of Mean Recognition Thresholds

Source	*df*	*MS*	*F*
Between *Ss*			
Anxiety (A)	1	223.73	
Set (S)	1	8302.17	
Frequency (F)	1	329643.96	19.45**
A × S	1	92841.90	5.48*
A × F	1	1151.33	
S × F	1	61065.64	3.60
A × S × F	1	914.41	
Error	72	16945.24	
Total	79		
Within *Ss*			
Words (W)	1	61525.20	58.11**
W × A	1	2513.81	2.37
W × S	1	31992.04	30.22**
W × F	1	47.79	
W × A × F	1	2939.51	2.78
W × A × S	1	677.24	
W × S × F	1	142.08	
W × A × S × F	1	464.11	
Error	72	1059.72	
Total	80		
Grand Total	159		

* Significant at .03 level.
** Significant at <.01 level.

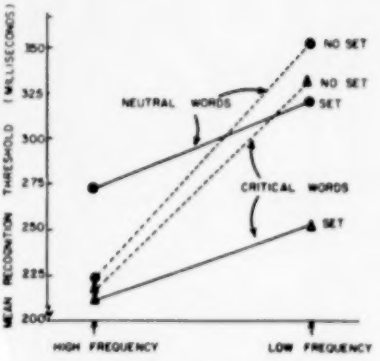

Fig. 1. Mean recognition thresholds for critical and neutral words as a function of set and frequency of usage of the stimuli.

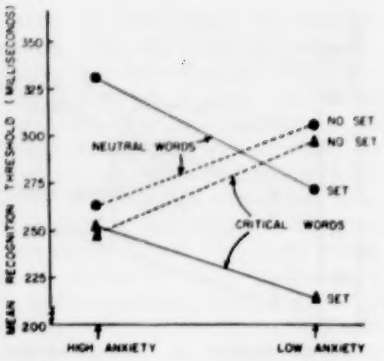

FIG. 2. Mean recognition thresholds for critical and neutral words as a function of set and anxiety level.

A graphical representation of the interaction of set and anxiety ($F = 5.48$, $P = .03$) is presented in Fig. 2. It should be noted that the effects of set for both critical and neutral words are differential as a function of anxiety level. This would be consistent with the failure to find any significant main effects of anxiety since the functions are, in fact, reversed. In the case of the critical words, set results in lower thresholds for the low-anxious *Ss* ($t = 2.73$, $P = .01$) while for the high-anxious *Ss* there is no significant difference in thresholds as a function of set. For the neutral words, there is no significant difference between set and no-set for the low-anxious *Ss*, while for the high-anxious *Ss* recognition thresholds were significantly lower ($t = 2.00$, $P = .05$) under the no-set condition. Apparently, set functioned in the expected manner with respect to the critical words for the low-anxious *Ss* while resulting in raised recognition thresholds for the high-anxious *Ss* in the case of the neutral words. In addition, when the recognition thresholds to critical words under the set condition are considered, there were no significant differences as a function of anxiety level although the thresholds to the neutral words under the set condition were significantly higher ($t = 2.38$; $P = .03$) for the high-anxious *Ss*.

The results of the set and anxiety interaction (for all stimulus words) may be generally summarized as indicating a trend toward higher thresholds for the high-anxious *Ss* under the set condition while lower thresholds result for the high-anxious *Ss* under the no-set condition.

DISCUSSION

Clearly, the effects of set were not uniform at different anxiety levels. In terms of our original expectations, set and frequency both were postulated as associative variables defining the difficulty of the perceptual task. That this assumption is justified in the case of frequency is indicated by the more or less consistent effects of this variable upon thresholds. However, in the case of set, the fact that thresholds were actually higher for the high-anxious *Ss* under this condition rather than lower as was originally expected would appear to take exception to the consideration of set as providing an increment to associative strength, at least for the high-anxious *Ss*. The consideration of set as an associative variable (as originally conceived) could apply to the performance of the low-anxious *Ss*, however.

It is apparent that the inconsistency of the functioning of set at different anxiety levels must suggest some additional process in operation here. The set and no-set conditions may be considered as two conditions differing in the degree of competitive response tendencies involved, being greater in the former than in the latter. That is, under the set condition, *S* is set to discriminate between set and non-set (critical and neutral) words whereas this condition necessarily does not prevail under the no-set condition. Looking at the performance of the high-anxious *Ss* under

the set condition, then, which involves competition of response tendencies, one sees that these *Ss* perform poorer (higher thresholds) than the low-anxious *Ss*. However, under the no-set condition (no competing response tendencies) the performance of the high-anxious *Ss* is superior (lower thresholds) to that of the low-anxious *Ss*. Such an interpretation of the results would tend to be consistent with the general findings of Spence and his associates (**8, 12, 18**) that in a competitive response-type situation, anxious *Ss* tend to perform poorer than non-anxious *Ss*, whereas they perform better in a simpler situation.

Therefore, within the present experiment, the set condition, paradoxically, would have to be considered the more difficult behavioral situation, while the no-set condition would be simpler from the standpoint of competing response tendencies. Obviously, such a conception of the properties of set is quite contradictory to the traditional view which would contend that competing response tendencies or conflicting hypotheses would necessarily be reduced by set. That such a competitive situation does develop is quite feasible though, especially in light of our results which might be said to indicate a differential effect of competitive response tendencies upon performance, the effects being less in the case of the low-anxious *Ss* and greater in the case of the high-anxious *Ss*.

If the performance of the high- and low-anxious *Ss* under the set condition for the two types of words is considered, it may be noted that the differences in thresholds for the neutral words is much greater than the differences for the critical words as a function of anxiety level. This would seem to indicate that it is the difference in the recognition of the neutral words which most markedly distinguishes the performance of these two groups. Put still another way, it would suggest that among the effects upon perception of increasing the levels of motivation is the impairment of recognition of stimuli which are not appropriate or are "incidental" to the set. In effect, the perception of these "incidental" stimuli is made less probable by increasing the level of motivation. An analogous finding in the field of learning has been obtained by Bruner, Matter, and Papanek (**1**) who demonstrated that by increasing levels of motivation, the range of cues to which the organism is attentive is reduced, thereby rendering less probable any learning incidental to the main task at hand. In our study, the performance of the high- and low-anxious *Ss* relative to the recognition of the neutral words would seem to reflect the operation of a similar mechanism involving impaired breadth of perceptual functioning with increasing levels of motivation.

Summary

An experiment was performed to test the hypothesis that perceptual recognition thresholds are a joint function of associative and drive strength. Eighty *Ss* selected on the basis of extreme scores on the Taylor A scale constituted high- and low-drive groups. Set of the *Ss* and frequency characteristics of the stimulus materials were manipulated as associative variables. The data were based upon tachistoscopic mean recognition thresholds for critical words (those congruent with set) and neutral words (those incongruent with set) and support the following conclusions:

1. The influence of Frequency, Words, and Words × Set were highly significant in the determination of recognition thresholds.
2. Set was found to interact with both frequency of the stimuli and anxiety of *Ss*.
3. The interaction of set and anxiety indicated that set facilitated the perceptual performance of the low-anxious *Ss* while impairing performance of the high-anxious *Ss*.
4. The greatest differences in performance between high and low-anxious *Ss* involved the imparied recognition of neutral words, or words incidental to the set by the high-anxious *Ss*.

Alternative interpretations of the properties of set under the present experimental conditions were discussed since details of the original hypothesis were not confirmed. The results, in general, were seen to be consistent with those of studies showing poorer performance for anxious *Ss* in a competitive response-type situation. Portions of the results were also interpreted from the standpoint of impaired

breadth of perceptual functioning associated with high levels of drive.

References

1. Bruner, J. S., Matter, J., & Papanek, M. L. Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.*, 1955, **62,** 1–10.
2. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45,** 120–125.
3. Freeman, J. T., & Engler, J. Perceptual recognition thresholds as a function of multiple and single set and frequency of usage of the stimulus material. *Percept. Mot. Skills*, 1955, **5,** 149–154.
4. Holtzman, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, **47,** 853–854.
5. Howes, D. H., & Solomon, R. L. Visual duration thresholds as a function of word probability. *J. exp. Psychol.*, 1951, **41,** 401–410.
6. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
7. McGinnies, E., Comer, P. B., & Lacey, O. L. Visual recognition thresholds as a function of word length and word frequency. *J. exp. Psychol.*, 1952, **44,** 61–64.
8. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45,** 91–96.
9. Postman, L. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation: a symposium.* Lincoln: Univ. of Nebraska Press, 1953.
10. Postman, L., & Bruner, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39,** 369–377.
11. Postman, L., & Crutchfield, R. S. The interaction of need, set and stimulus structure in a cognitive task. *Amer. J. Psychol.*, 1952, **65,** 196–217.
12. Ramond, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46,** 120–124.
13. Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43,** 195–201.
14. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45,** 116–119.
15. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.
16. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41,** 81–92.
17. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.
18. Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44,** 61–64.
19. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Columbia Univ. Press, 1944.
20. Welch, L., & Kubis, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23,** 83–91.
21. Wispé, L. G., & Drambarean, N. C. Physiological need, word frequency, and visual duration thresholds. *J. exp. Psychol.*, 1953, **46,** 25–31.

(Received June 17, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

RESPONSE GENERALIZATION AS A FUNCTION OF INTRATASK RESPONSE SIMILARITY [1]

MERRILL E. NOBLE [2] AND HARRY P. BAHRICK [3]

The Ohio State University *Ohio Wesleyan University*

It has frequently been pointed out (**7,** p. 309) that a progressive restriction in the variability of responses characterizes certain aspects of the learning process. Only rarely, however, do we find a detailed description of response distributions to specify the nature of these changes. Investigators are usually careful in devising procedures for the detailed control and description of independent variables, but they are often content to employ relatively gross response indices as dependent variables (**3**). Examples of such gross response measures are the all-or-none records of key-pressing or the right-wrong scoring of verbal responses. Perhaps this practice reflects the belief that the most immediate problems of the behavioral sciences lie in the specification of important antecedent conditions of behavior, and that emphasis upon a more analytical examination of response changes must await the refinement of theory.

It can be argued, however, that even at their present stage of development behavioral theories are not well tested because of the relative lack of precision used in the measurement of responses. This view suggests that studies employing a more analytic description of responses might permit a clarification or resolution of some theoretical controversies. One area in which such an approach might be especially fruitful concerns facilitative and interference effects due to response similarity. Several somewhat different viewpoints are held regarding the effects of intratask response similarity (**3, 8, 9, 13**). These viewpoints are based for the most part upon the results of studies in which gross response indices are used, and in which response similarity is varied simply along an ordered scale, but not specified along a continuous equal-interval scale.

The present paper reports an investigation in which continuous response distributions were obtained for two versions of a task. The two versions represented different intratask response similarities. Response similarity was itself defined by reference to an interval distance along a continuous scale. This made it possible to specify the variance of the response distributions to each stimulus, the constant errors, and the average amount of information transmitted per response. This last measure also reflects the amount of overlap among response distributions.

PART I

Method

The study was conducted in two parts. In Part I a psychophysical scale of force similarity was constructed. In Part II response distributions were obtained under two conditions of similarity among the required responses.

Apparatus.—The apparatus consisted of a semirigid steel control stick, mounted as shown in Fig. 1, and an indicator sensitive to .001 in. of movement. The indicator was placed in such a

[1] This study was conducted in the Laboratory of Aviation Psychology of The Ohio State University, through support of Contract No. AF 33(038)-10528 between the Air Force Personnel and Training Research Center and The Ohio State University Research Foundation.

[2] Now at Kansas State College.

[3] The authors are indebted to Drs. Paul M. Fitts, David Bakan, and Benton J. Underwood for many constructive criticisms.

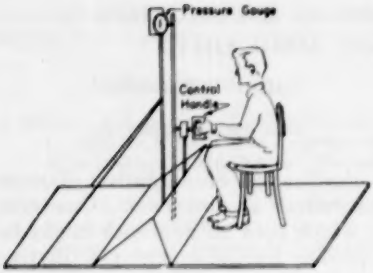

FIG. 1. Schematic drawing of the pressure control.

position that each mark on the scale represented ¼-lb. force exerted on a handle by *S*. This system was approximately isometric, with .1 in. of movement resulting when a force of 50 lb. was exerted on the stick handle.

Subjects.—Ten right-handed male students at Ohio Wesleyan University were employed on an hourly basis. Each *S* served under each of the 10 conditions of force.

Forces.—Ten forces, 1, 2, 3, 5, 10, 15, 20, 30, 35, and 40 lb., were used.

Procedure.—Each *S* made a total of 80 responses to a single one of the 10 forces during each of ten ½-hr. sessions. The 10 sessions were separated from one another by at least 24 hr. The sequence of forces required of each *S* in successive sessions was counterbalanced by means of a random latin square. The *S* was blindfolded and instructed to push a stick handle slowly when *E* said "now" until he thought the desired force was attained, and to then release the handle and relax until the next response. The *S* was told he would have to learn to exert the desired amount of force by trial and error. After each response *E* recorded the amount of force to the nearest ¼ lb. and told *S* his error in ¼-lb. units., e.g., if the desired force was 10 lb. and *S* exerted 11-lb. force he was told "four units over." There was a 15-sec. rest between responses.

Results

Learning curves plotted for successive blocks of 10 responses indicated that there was no appreciable learning beyond 10 trials on any stimulus. Accordingly, means and *SD*'s of the distribution of responses were determined for each *S* for each force for Trials 11–80. Inspection of these data indicated that constant errors were small and were normally distributed among *S*s for each force. Figure 2 shows an empirical gradient of response generalization for a stimulus of 10 lb. The responses are pooled for all *S*s. Equal discriminability units on the abscissa are obtained

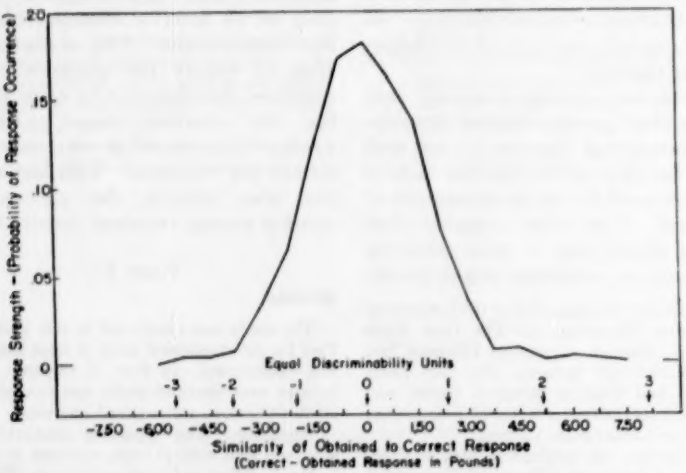

FIG. 2. An empirical gradient of response generalization for a response of 10 lb.

from Equation 1 by a procedure described below.

The average *SD* for each distribution of responses was computed as the square root of the within-groups mean square for each of the 10 forces. The 70 responses of each *S* to each stimulus were treated as a group, so that the *SD*'s include only variations within *S*s, but not variation between *S*s. Figure 3 shows the relation between the required force and the ratio of the average *SD* to the required force. Also shown in Fig. 3 are similar data from a study by Jenkins (**5**) in which *S*s attempted to reproduce a given experienced force by pushing against a semirigid control stick. Examination of these data indicates that the Jenkins study and the present one yielded functions of a similar type although Jenkins' *S*s had less variance than *S*s of the present study. The relation between the average *SD* and the amount of required force is shown in Fig. 4. Inspection indicated that this relation approximates a straight-line function expressed by the equation

$$SD = .1_p + .54 \qquad (1)$$

where *SD* is the square root of the within-group mean square in pounds and p is the force in pounds.

A scale of force similarity was then determined. Forces were selected in such a way that successive forces were separated by the square root of the sum of squares of their respective average *SD*'s. The first force was arbitrarily chosen to be 1 lb. The remaining forces were determined by solving appropriate quadratic equations, using as constants the values obtained from Equation 1. The resulting scale approximates an equal-interval scale, since there should be approximately an equal percentage of overlap of response distributions for all adjacent forces. The values obtained for the scale are shown in Table 1.

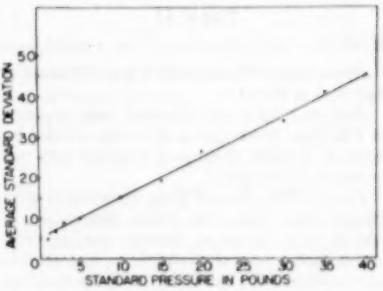

FIG. 4. Average *SD* as a function of pressure.

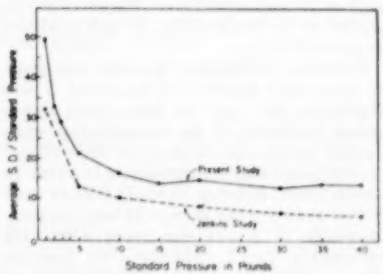

FIG. 3. Average *SD* per required pressure as a function of pressure.

TABLE 1

THE SEVEN REQUIRED FORCES

Scale Values from Part I	Force in Pounds												
	1.00	1.98	3.10	4.40	5.89	7.63	9.63	11.89	14.53	17.63	21.13	25.12	29.78
Condition N				4.50	6.00	7.75	9.75	12.00	14.50	17.75			
Condition W	1.00		3.00		6.00		9.75		14.50		21.25		29.75

Part II

Method

Apparatus.—The apparatus was the same as that used in Part I.

Subjects.—Fifty right-handed male students at The Ohio State University were divided by means of a table of random numbers into two groups of 25 *Ss* each.

Forces.—Two sets of seven forces each were selected from the scale values determined in Part I. For Cond. N (narrow spacing) successive forces were separated from one another by one scale unit; for Cond. W (wide spacing) successive forces were separated from one another by two scale units. Three forces, including the median force for each set, were common to both conditions. The seven forces, rounded to the nearest ¼ lb., for each condition are given in Table 1.

Procedure.—Preliminary practice was given on the central force (9.75 lb.). The *S* was blindfolded and told to exert force slowly against the handle of the semirigid stick when *E* called out the stimulus number "39." The *S* was instructed that the stimulus represented an arbitrary unit, and that he would have to learn by trial and error the amount of force required. Knowledge of results was given after each response as in Part I. The preliminary practice continued until *S* responded with a force between 8¼–11 lb. Following this, each *S* made a total of 280 responses in a single session, 40 responses with each of the seven stimuli of the condition to which he had been assigned. Within this restriction the 280 responses were randomized by using three different random orders of stimuli for each group of *Ss*. There was an 8-sec. rest between responses.

For both conditions the stimulus for each response was the amount of force to be exerted, called out by *E* in ¼-lb. units. Thus, if the required force was 6 lb., *E* called out "24"; for a force of 9.75 lb., *E* said "39"; etc. The *S* was blindfolded and, as before, told only that the stimuli represented arbitrary units. Knowledge of results was given after each response as in Part I.

It should be noted that the stimuli for Cond. N and W were different, i.e., for each condition the stimulus was a number representing the particular amount of force to be exerted in ¼-lb. units. It is possible that this difference in the stimuli of the two conditions could affect the results. That is to say, even though for both conditions the stimuli were highly discriminable from one another, it is possible that *S* adjusted the error he would tolerate to the amount of separation between adjacent stimuli instead of having a constant set to respond as closely as possible to the required force. Since for Cond. W the required forces were further apart than for Cond. N, a flexible error tolerance would tend to result in larger variances for Cond. W than for Cond. N. Consequently, a group of 15 right-handed *Ss* was added at a later date in order to test this effect. The stimuli for this control condition (Cond. C) were identical to those of Cond. W; the required responses were those of Cond. N. Knowledge of results were given as in Cond. N, i.e., the number of ¼-lb. units over or under the forces desired in Cond. N. Otherwise, the procedure was the same as before.

Results

For both conditions the mean constant error (*CE*) and the square root of the within-group mean square were computed for successive 5-trial blocks. The *SD*'s again include only variance within *Ss*, not variance between *Ss*. These data are presented in Table 2. Inspection of the data shows that learning resulted primarily in a reduction in the magnitude of the *CE*'s, and to a much lesser extent a reduction in the variability around these *CE*'s. Early in training there is a pronounced range effect for both groups, i.e., a tendency to respond with too much force to the lowest desired responses, and with too little force to the highest desired responses. This tendency has been observed by other investigators (**4**), but it is noteworthy here that this effect is transitory, and with continued practice the direction of the *CE* tends to become positive even for the higher forces, i.e., *Ss* tend to exert more than the required amount of force. Of greatest interest are the data for the three stimuli common to both groups: "24," "39," and "58" quarter-pounds. Inspection of Table 2 indicates that little or no learning took place after the initial 20 responses to these stimuli for both groups, and during the final 20 responses to these stimuli the *CE*'s

TABLE 2

MEAN RESPONSE AND *SD* FOR SUCCESSIVE BLOCKS OF FIVE RESPONSES

Trials	Condition N—Desired Force (In Quarter-Pounds)													
	18		24		31		39		48		58		71	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1–5	26.17	6.30	30.68	5.80	36.00	7.00	38.66	6.69	44.09	9.01	50.63	8.98	62.74	11.77
6–10	22.33	4.49	29.13	5.93	35.13	5.66	40.30	5.41	47.54	5.89	54.62	7.71	66.58	11.79
11–15	24.17	6.85	28.80	5.88	35.72	5.67	39.57	5.89	46.22	6.57	53.22	8.66	69.10	10.76
16–20	22.29	4.15	29.09	5.80	35.29	5.79	40.71	5.84	46.58	6.05	53.74	9.24	70.32	7.10
21–25	20.46	5.77	26.99	5.79	34.42	6.12	40.35	6.43	46.39	7.05	54.23	7.35	71.54	9.18
26–30	20.75	4.96	27.02	5.54	36.06	6.67	40.83	5.16	49.86	6.42	58.27	9.00	72.94	9.07
31–35	21.07	5.44	26.66	5.81	36.54	8.65	41.72	6.83	49.63	6.97	57.93	10.21	73.66	12.57
36–40	21.54	5.24	26.93	5.45	34.99	5.45	42.58	5.41	49.32	6.88	56.92	8.30	72.86	9.07
	Condition W—Desired Force (In Quarter-Pounds)													
	4		12		24		39		58		85		119	
1–5	12.62	5.57	22.46	10.10	33.54	9.87	39.97	10.62	49.25	11.95	65.66	14.92	96.58	22.24
6–10	9.42	4.05	18.81	5.61	33.73	9.61	46.06	10.92	61.82	12.56	75.39	14.34	110.25	18.96
11–15	9.03	2.46	20.14	6.87	30.44	8.60	44.40	11.29	60.19	11.04	79.49	14.69	115.87	17.72
16–20	6.81	2.84	16.43	5.04	31.46	8.38	46.86	13.76	60.47	13.85	81.62	9.76	114.68	17.16
21–25	6.98	3.17	18.10	6.10	29.76	8.13	42.42	9.64	63.65	12.76	80.62	12.91	115.74	16.29
26–30	5.19	2.61	14.81	5.98	29.57	6.83	42.27	10.80	60.02	11.27	78.23	14.55	116.30	17.76
31–35	6.74	3.85	16.38	8.46	27.89	7.54	44.19	11.82	64.55	11.22	80.34	13.33	119.36	15.02
36–40	5.88	2.41	16.26	5.99	29.77	8.07	46.40	8.74	63.90	10.97	83.50	10.73	119.77	16.24

and within-group variances are relatively stable. The *CE*'s and mean within-group variance for these last 20 responses to the three common stimuli were computed for both groups and are shown in Table 3. *T* tests show that the mean within-group variance is significantly ($p < .01$) larger for Cond. W than for Cond. N for all three forces. Variability of variance among the 25 *S*s of each group was used as a basis for the error terms. Condition W also results in significantly larger constant errors ($p < .01$) than Cond. N.

The *CE*'s and mean within-group variances for the last 20 responses to each stimulus were also computed for the control group. None of the *CE*'s or variances differs significantly ($p > .05$) from those obtained for Cond. N.

The average amount of response information (I_r), response equivocation (E_r), and information transmitted (I_t) was computed for Cond. N and

TABLE 3

COMPARISON OF THE THREE RESPONSE DISTRIBUTIONS COMMON TO CONDITIONS N AND W, BASED ON THE LAST 20 RESPONSES TO EACH FORCE

	Force (In Quarter-Pounds)		
	24	39	58
Cond. N			
Mean	26.90	41.37	56.84
SD	1.85	2.06	3.70
Variance	26.59	30.58	66.47
SD	11.68	14.91	24.61
Cond. W			
Mean	29.25	43.82	63.03
SD	2.23	2.44	4.82
Variance	49.42	91.03	114.86
SD	12.21	21.33	38.57

TABLE 4

INFORMATIONAL ANALYSIS OF THE DATA

Cond.	I_r (Bits)	E_r (Bits)	I_t (Bits)
N	6.010	4.920	1.090
W	6.775	5.188	1.587

W. For these computations a 7-column (representing the seven required forces) by 165-row (representing forces actually exerted, in ¼-lb. units) matrix was used. The data for all *S*s in a given condition were entered in the matrix, i.e., the distributions of the last 20 responses to each of the seven required forces were pooled for all *S*s. This procedure, which results in a slightly inflated estimate of E_r because of pooling of separate *S*'s response distributions, was followed because of the limited number of responses per *S* and because it permitted a different comparison of the two conditions based upon group performance and ignoring individuals. The results of this analysis are shown in Table 4.

Reference to Table 4 shows that the greatest amount of information was transmitted under Cond. W. Thus, even though Cond. W resulted in the larger *CE*'s and the larger within-groups variance per individual response distribution, this result indicates that there was less overlap between adjacent response distributions for this condition than for Cond. N.

DISCUSSION

The response distributions obtained in Parts I and II are interpreted by the authors as empirical gradients of response generalization. However, since our experimental situation is not the type traditionally used for the observation of response generalization, it is necessary to indicate the ways in which the procedures followed here conform to the essential operations that define response generalization.

Response generalization is said to exist when "A given stimulus condition, once connected with a particular response will also elicit other responses which are related in some way to the first or trained response" (**6**, p. 326). In the traditional method used for the study of response generalization, a particular stimulus is presented a number of times during a training series and one specific response to this stimulus is reinforced (the reinforcement may involve merely the presentation of the correct response on a memory drum). Other ("incorrect") responses to the stimulus usually occur during this training period, but these responses supposedly are not reinforced. At some arbitrary point in time the training is terminated and measurements are made of the frequency of responses similar to the one reinforced during training (**13, 14**).

In comparing this traditional procedure with the one used in the present study, it is seen that the present study lacked the initial controlled training period during which positive reinforcement is given to a specific stimulus-response combination. Instead, in the present study a large number of similar responses occurred to a given stimulus on different trials. All of these responses may seem to have been equally reinforced! It should be noted with respect to the typical procedure, however, that a variety of slightly different responses will occur during all training situations, though the response deviations that are observed may be only relatively large ones. That is to say, the response units usually employed in discrete S–R tasks are relatively gross, and indeed are often of an all-or-none character. Thus, a specific key is, or is not, pressed; a specific word is, or is not, spoken. Minor variability, such as in the force applied to depress a key, is usually not observed, and therefore a variety of responses will be reinforced. Only gross errors are recorded in such situations and these are not reinforced. Such errors would occupy the tails of the response

distributions in the present study, or they would occur through secondary or mediated generalization.

Examination of the data of the present study provides some evidence also that not all responses are equally reinforced, since there was a progressive change toward the response designated as correct, at least during the early part of training. When an incorrect response occurred (e.g., a response of "55" to a stimulus of "39") *E*'s statement of "16 over" did not directly strengthen this response; instead, *S* would tend to use less force on the next such trial, thus bringing his response closer to the desired value. Nevertheless, the most important feature of these data is that a variety of responses were produced to each stimulus throughout training, despite differential reinforcements. The resulting gradient of responses can be attributed to a partial failure of discrimination between the specific correct response and the other similar responses. If generalization and discrimination are considered to be at opposite poles of a continuum, this is tantamount to stating that generalization occurred, and in an amount inversely related to the similarity between the correct and the generalized response. With this view, response generalization can be observed during the acquisition training of any response by noting the occurrence of existing response variations. This has already been suggested by Brogden, (**2**, p. 590) and has formed the basis of an earlier study by Arnold (**1**).

Reference to the data of discrete S–R studies suggests much greater differences between the correct and generalized responses than were observed in the present study. A precise comparison is not possible, of course, without comparable continuous scales of response similarity. It seems likely, however, that the generalized responses of discrete S–R tasks often reflect instances of secondary or mediated generalization, while the present data and those obtained by Arnold exemplify primary response generalization.

In this connection, it is of interest to compare the other results of this study with those of studies using discrete S–R tasks. The results here show that practice in making the more similar responses (Cond. N) resulted in smaller variances and *CE*'s for the response distributions than was the case when the less similar responses (Cond. W) were made. Furthermore, a comparison of the *SD*'s obtained in Part II with those predicted from Equation 1 from the data of Part I, shows that the predicted values for Cond. N closely approach the obtained values for five of the seven response distributions (for the two lowest forces the obtained *SD*'s were larger than predicted). This was not the case for Cond. W, however. There, the obtained *SD*'s were all substantially larger than the predicted values. The failure to find significant differences between the variances or the *CE*'s for Cond. N and C suggests that these results cannot be attributed to an increase in amount of separation among the already highly discriminable stimuli of the tasks. The informational analyses showed that the amount of overlap among response distributions was greatest for the condition having the greater response similarity, despite the finding that an increase in intratask response similarity led to decreased response variability.

As noted earlier, "minor" variability usually is ignored in discrete S–R tasks, and only such gross errors as calling out an entirely different word or pressing another button are recorded. Such studies have shown that when stimuli are held constant an increase in intratask response similarity results in an increased number of intrusion errors (**10, 11, 12**). These findings are often accounted for by means of the concept of response generalization in one form or another (**3**). The generalized response strength is said to be inversely related to the degree of similarity between the generalized and the reinforced response. If this assumption is applied to the data here it would suggest that the overlap of response distributions should be a simple direct function of the similarity among the required responses of the task. The present results, however, show that the distribution of responses around

designated response points becomes less variable as response similarity is increased. If this result is a general one, it suggests that intrusion errors due to primary response generalization are not a simple function of response similarity, since the shape of the generalization gradients themselves is also affected by the degree of response similarity. In order to determine whether the present results have generality, it will, of course, be necessary to develop other tasks in which response similarity can be specified along a continuous scale.

The results obtained from the information analysis have a bearing upon the problem of coding of responses. They suggest that for optimum coding the range of responses should be extended so that the scale interval between response categories is maximized. The informational analyses suggest that even with optimum spacing of pressure responses between 1 and 30 lb., only about three absolutely discriminable categories of responses are possible if the responses of several *S*s are placed in a single distribution. Information transmission for the responses of single individuals would doubtless be greater.

Summary

This study points out the advantages of obtaining distributions of responses permitted to vary along a scaled continuum in order to determine the effects of response similarity upon response generalization.

The *S*s were trained to exert different amounts of force on a semirigid control in response to number stimuli. Intratask similarity was varied by employing two versions of the task, differing in the degree of separation between required adjacent force responses. Responses were measured on a continuous scale and response distributions were obtained for several response points under both conditions of response similarity. These distributions are interpreted as empirical gradients of response generalization.

It is shown that generalization of individual responses is significantly smaller for the condition of greater intratask response similarity. This condition results in steeper generalization gradients. Despite this effect adjacent response distributions show more overlap when the required responses are spaced more closely. The procedure and results of this study are compared with those of discrete S–R tasks investigating the effects of response similarity.

References

1. Arnold, W. J. An exploratory investigation of primary response generalization. *J. comp. Psychol.*, 1945, **38**, 87–102.
2. Brogden, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Ch. 10.
3. Duncan, C. P. Development of response generalization gradients. *J. exp. Psychol.*, 1955, **50**, 26–30.
4. Ellson, D. G., & Wheeler, L., Jr. *The range effect*. Dayton, O.: USAF Air Materiel Command, Wright-Patterson Air Force Base, 1949. (*AF Tech. Rep.* No. 5813).
5. Jenkins, W. O. The discrimination and reproduction of motor adjustment with various types of aircraft controls. *Amer. J. Psychol.*, 1947, **60**, 397–406.
6. McGeoch, J. A., & Irion, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
7. Skinner, B. F. *The behavior of organisms: an experimental analysis*. New York: Appleton-Century, 1938.
8. Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.
9. Underwood, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *J. exp. Psychol.*, 1951, **42**, 44–53.
10. Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired adjective lists with two levels of intralist similarity. *J. exp. Psychol.*, 1951, **42**, 153–161.
11. Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142.
12. Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149.
13. Underwood, B. J., & Hughes, R. H. Gradients of generalized verbal responses. *Amer. J. Psychol.*, 1950, **68**, 422–430.
14. Wickens, D. D. The transference of conditioned excitation and conditioned inhibition from one muscle group to the antagonistic muscle group. *J. exp. Psychol.*, 1938, **22**, 101–123.

(Received June 9, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

A FAILURE TO FIND SECOND-ORDER SEMANTIC GENERALIZATION

IRVING MALTZMAN AND LLOYD O. BROOKS

University of California, Los Angeles

Probably the most extensive series of studies on semantic generalization have been conducted by Cofer, Foley, and their associates. Two of their studies provide what appears to be the most unequivocal evidence for the hypothesis. One study reported the occurrence of semantic generalization between synonyms, synonyms of synonyms (second-order semantic generalization) and between two different lists of homophones (**2**). Another study purported to show semantic generalization between antonyms (**1**). A verbal transfer method was utilized in both experiments.

The *Ss* were first given a buffer or practice list followed by a written recall and then an equating list and its recall test. They were then distributed among five different groups, each receiving four presentations of a reinforcement list. Reinforcement lists practiced by the experimental groups were related semantically to a subsequent test list. The latter was presented only once and followed by a written recall. Two groups received different lists of homophones. The other experimental conditions were synonyms and second-order synonyms. A control group received a reinforcement list of words unrelated to the test list. The *Ss* were told that they would not be asked to recall the reinforcement list. Since *Ss* were tested for recall of the first two lists, however, the effectiveness of these instructions for the reinforcement list is questionable.

In terms of mean test words recalled, only one of the homophone groups did significantly better than the control group. However, in terms of medians all four experimental groups differed significantly from the control group. But our application of Shen's test (**7**) to the data indicated that for the control group the mean would be a more reliable measure than the median. In view of this finding, and the lack of agreement between means and medians, the reliability of the results of this experiment is highly questionable.

The experiment on semantic generalization between antonyms (**1**) followed essentially the same procedure as the first study. All *Ss* received a buffer and an equating list. An experimental group received a reinforcement list of antonyms of the test words while a control group received words unrelated to the test list. An important variation from the first experiment was the addition of a second control group which did not receive a reinforcement list, but was given the test list immediately after the written recall of the equating list.

Semantic generalization between antonyms was reported to have occurred even though the mean difference between the experimental group and the first control group was significant at only the 10% level (two-tailed test). Mean recall of the second control group was compared only with the first control group from which it did not differ significantly. Inspection indicates that the second control group likewise did not differ significantly from the experimental group. Differences among the medians of the three groups were even smaller than the mean differences, but their reliabilities were not reported.

The absence of a significant difference between the experimental group and the second control group contradicts the hypothesis that semantic generalization occurred in the former group. If repetition of the antonym reinforcement list produced increments in habit strength which generalized to the test list, the test words should have been recalled more readily by the experimental than by the control group that did not receive any reinforcement list. The absence of any such difference implies that if a condition comparable to the second control group had been run in the first experiment, evidence clearly contrary to the hypothesis of semantic generalization would have been obtained here as well.

Despite the questionable nature of the differences obtained between the experimental and control groups in these experiments, Osgood (**3, 4**) has offered an hypothesis which he proposes as an alternative to the semantic generalization interpretation of the data. He argues that if experimental *Ss* recognize the presence of a semantic relationship between the reinforcement and test lists, they would be able to reconstruct the test list, at least in part, from

their recall of the words of the reinforcement list. A control group that receives unrelated reinforcement words could not reconstruct the test list, and therefore would have a poorer recall score. It would be easier to reconstruct the test-list words from recall of their homophones than from synonyms, because of the greater number of alternative responses in the latter case. This would account for the greater apparent generalization between homophones than between synonyms found with the verbal transfer method. In contrast, greater generalization between synonyms has been found by the conditioning procedure (**5, 6**).

Although Osgood's hypothesis gives a satisfactory alternative explanation of the superior recall of the homophone and synonym conditions as compared with the control condition, provided that we assume that these differences are reliable, it cannot account for the superior recall of the second-order synonym condition. Reconstruction of the test list from synonyms of their synonyms seems unlikely, because there is no apparent semantic relationship between the reinforcement and test words, for example, *killed* and *sent*. Even if this relationship were known, there are such a large number of alternative responses possible, that it seems improbable that reconstruction of the test list could occur. If reconstruction is not possible, then increased recall of the test list as a consequence of practice on synonyms of their synonyms would have to be attributed to the effects of semantic generalization.

Second-order synonyms were therefore chosen for study in the present experiment because positive results could be interpreted more readily than in the other semantic conditions. To further ensure that reconstruction of the test list could not occur, the present experiments were designed to reduce the possibility of *S*s' discovering the relationship between reinforcement and test words. The first experiment to be reported here was also designed to determine whether second-order semantic generalization, if demonstrable, varies as a function of the number of reinforcements.

Experiment I

Method

Subjects.—The *S*s were 120 students obtained from introductory psychology classes. There were 20 *S*s randomly assigned in pairs to each of the six different conditions.

Materials.—The reinforcement and test words employed were obtained from the lists given by Foley and Cofer (**2**). Each word was printed on a 2 × 2-in. slide and projected on a beaded screen by means of a Spindler and Sauppe Selectroslide projector.

Procedure.—All *S*s received a reinforcement series of 10 different slides followed after a 1-min. rest period by a test series of 20 slides. Each reinforcement series presented to the experimental groups contained a word which was a synonym of a synonym of a critical word in the test list. Three experimental groups differing only in the number of repetitions of the reinforcement series were employed. These groups received 4, 8, and 20, repetitions, respectively. Foley and Cofer's (**2**) groups received 4 repetitions of the reinforcement list. Three control groups receiving 4, 8, and 20 repetitions of the reinforcement series were also employed. The reinforcement words for these groups were not related in any obvious manner to the critical or the noncritical test words. Duplicate slides were made of each word in the experimental and control practice lists. It was therefore possible to lay out a prearranged random order of slides for each repetition of the practice lists. This procedure eliminated the need for reshuffling the slides between repetitions in order to randomize their serial positions.

The test list given to the experimental and control groups contained the 10 critical words employed by Foley and Cofer (**2**). Each of these words had a synonym of its synonym in the reinforcement list presented to the experimental groups. In addition, 10 noncritical words were randomly interspersed among the critical test words. The noncritical words were added to the test list to reduce the possibility of experimental *S*s recognizing a semantic relationship between the reinforcement and test lists. The noncritical words were Christian names selected from Foley and Cofer's (**2**) equating list.

The *S*s were run in pairs in all six conditions. Prior to the presentation of the reinforcement series they were told that the experiment was an investigation of the speed of perceptual recognition. They were asked to call out the word as soon as it flashed on the screen, and that their speed of response would be recorded. Prior to the presentation of the test list *S*s were instructed not to call out, but to think of the meaning of each word. They were also told that they would be tested on their retention of this series.

Following one presentation of the test series, *S*s were given a sheet of paper and were asked to record as many of the test words as possible.

TABLE 1

MEAN RECALL OF REINFORCEMENT WORDS

	Control Groups			Experimental Groups		
	C_4	C_8	C_{20}	X_4	X_8	X_{20}
Mean	3.75	3.80	6.35	4.85	4.85	5.90
SD	4.27	4.26	6.73	5.22	5.54	6.22

There was no time limit. They were then asked for a written recall of the words from the reinforcement series. The *Ss* were questioned as to their interpretation of the purpose of the experiment and as to the presence of any relationship or similarity between the practice and test lists.

Results—Table 1 shows the mean number of reinforcement words recalled by the experimental and control groups. Hartley's test (**8**) indicated that the variances were homogeneous. An analysis of variance resulted in a significant F for practice effects, $F = 12.99$, for 2 and 114 df, $P < .01$. The differences between recall of the experimental and control reinforcement lists and the interaction between amount of practice and the kind of list were not significant.

Table 2 shows the mean recall of critical and noncritical test words obtained by the experimental and control groups. Hartley's test indicated that the variances for the recall of noncritical words by the different groups were homogeneous. None of the F ratios for the main effects or their interaction reached an acceptable level of confidence.

Harley's test indicated that the variances for recall of the critical test words were homogeneous. The analysis of variance for the critical test words likewise failed to indicate any significant main effects or interaction. The F for treatments was 1.05, $df = 1$ and 114. The F for practice level was .74, $df = 2$ and 114. These results indicate the absence of second-order semantic generalization effects.

None of the experimental *Ss* had any knowledge of the semantic relationship existing between the reinforcement words and the critical test words. However, a considerable number of experimental and control *Ss* assumed that there was some form of association or relationship between words of the reinforcement list and the noncritical words of the test list, the Christian names. The failure to find a significant treatment or practice effect in the recall of the test list may, therefore, have been due to competing responses aroused by the set to look for associations between reinforcement words and the noncritical test words. Another possible source of interfering effects may have been the competitive nature of the practice session.

EXPERIMENT II

The purpose of the second experiment was to determine whether different results would be obtained when the test list contained only critical words, as was the case in the Foley and Cofer study (**2**), and when

TABLE 2

MEAN RECALL OF CRITICAL AND NONCRITICAL TEST WORDS

	Control Groups			Experimental Groups		
	C_4	C_8	C_{20}	X_4	X_8	X_{20}
Critical Test Words						
Mean	3.65	4.30	4.30	4.35	4.25	4.45
SD	4.00	4.58	4.75	4.62	4.65	4.70
Noncritical Test Words						
Mean	4.90	4.50	5.70	5.00	4.85	4.70
SD	5.17	4.79	6.03	5.35	5.26	5.09

TABLE 3

MEAN RECALL OF TEST SERIES

	C_5	C_4	X_4
Mean	6.00	5.60	5.53
SD	6.31	6.02	5.87

Ss were run individually. A second control group was added in this experiment in order to determine whether the reinforcement lists were producing proactive inhibition for recall of the test list.

Method

Subjects.—Forty-five students from introductory psychology classes were employed in this experiment, 15 in each of the two control groups and the one experimental group.

Procedure.—The first control group (C_4) and the experimental group (X_4) were given four presentations of their respective reinforcement lists before the presentation of the 10 critical test words. The second control group (C_5) was given the test list without prior exposure to a reinforcement list. The *Ss* were run individually. In all other details the procedure was the same as in the previous experiment.

Results.—Table 3 shows the mean recall scores for the test list obtained by the three groups. The variances were homogeneous and the *F* ratio obtained from the simple analysis of variance was less than unity.

These results indicate that the failure to find a difference in Exp. I between the experimental and control groups in their recall of the test list was not due to the presence of noncritical words in the test list, or the competitive social nature of the experimental situation.

DISCUSSION

It may be argued, perhaps, that our failure to find second-order semantic generalization was due to the low level of motivation for learning the reinforcement list. The training conditions of these experiments approximate those of incidental learning. The *Ss* in Cofer and Foley's experiments, perhaps inadvertently, were more highly motivated to learn the reinforcement list due to their experience of recall tests on the two prior lists. Recall of the reinforcement list in our first experiment, however, showed a significant increase in retention with increased training despite the relative lack of motivation to learn. Since the words of the reinforcement list acquired increments in habit strength, as indicated by the superior recall of the test list of the experimental groups receiving a large number of repetitions of the reinforcement list, generalization to synonyms of their synonyms should have occurred.

The previous analysis of Cofer and Foley's experiments on semantic generalization and the results of the present experiments indicate that there is no sound basis for assuming that semantic generalization as manifested by verbal transfer occurs between homophones, synonyms, antonyms, or second-order synonyms. However, the possibility still remains that further increases in motivational level and number of reinforcements might produce second-order semantic generalization, and other methods may very well demonstrate these different kinds of semantic generalization (**5, 6**).

SUMMARY

Analysis of previous experiments on semantic generalization by Cofer and Foley using the verbal transfer method indicated that the evidence for semantic generalization between homophones, antonyms, synonyms, and second-order synonyms lacks an acceptable degree of reliability. Apparent generalization among semantic relationships other than second-order synonyms may be interpreted in an alternative manner by Osgood's reconstruction hypothesis. As a consequence, two experiments employing the method of verbal transfer were conducted in an attempt to obtain second-order semantic generalization. In the first experiment, experimental groups received varying amounts of practice on words which were synonyms of test words. Control groups received varying amounts of practice on words unrelated to the test words. Superior recall of

the critical test words by the experimental groups would indicate the occurrence of generalization. No difference between experimental and control groups was obtained in their recall of the test words. A second experiment designed to eliminate possible uncontrolled variables interfering with verbal transfer likewise failed to provide any evidence of second-order semantic generalization.

References

1. Cofer, C. N., Janis, M. G., & Rowell, M. M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32,** 266–269.
2. Foley, J. P., & Cofer, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32,** 169–175.
3. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49,** 197–237.
4. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univ. Press, 1953.
5. Razran, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90,** 89–90.
6. Riess, B. F. Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, **26,** 238–240.
7. Shen, E. Note on the sampling error of the median. *J. educ. Psychol.*, 1935, **26,** 154–156.
8. Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.

(Received May 27, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

TASK INSTRUCTIONS FOR ANAGRAMS FOLLOWING DIFFERENT TASK INSTRUCTIONS AND TRAINING[1]

IRVING MALTZMAN, EUGENE EISMAN,[2] LLOYD O. BROOKS, AND WILLIAM M. SMITH

University of California at Los Angeles

Previous studies by Rees and Israel (4) and by Maltzman and Morrisett (3) have demonstrated that task instructions may increase the disposition or set for a given class of solutions to anagrams possessing multiple solutions. The purpose of the present experiment was to determine whether task instructions would still have the effect of inducing a given set when prior instructions and training have induced a different set.

METHOD

Subjects.—Ninety-six volunteer *Ss* from introductory psychology classes served in the experiment. There were 12 *Ss* in each of the eight experimental subgroups. Each of the four *Es* conducted the experiment with a proportionate number of *Ss* in each condition.

Procedure.—All eight subgroups first received a training series of 20 five-letter anagrams each typed in lower case on 3 × 5-in. white index cards. There was only one solution for each of these anagrams. Half of the 96 *Ss* received training anagrams that could be solved by a word referring to an aspect of eating or food. For example, the anagram "ocnba" is solvable as "bacon." The anagrams were in no consistent letter order. Prior to the start of the training series, half of these 48 *Ss* received instructions to look for words referring to eating or different foods. This is the eating-instructions (EI) training group. The other 24 *Ss* received no special task instructions, thus forming the eating (E) training group.

The remaining 48 *Ss* also received training anagrams with only one possible solution. It was always in the same letter order, 32145. The third letter in the word was the first in the anagram, and so on. For example, the anagram "tamch" is solvable as "match." These words were not related to eating or any other category in any obvious consistent fashion. Prior to the start of the training series, half of these 48 *Ss* were given instructions to look for a given letter order in the anagrams. They formed the order instructions (OI) training group. The remaining 24 *Ss* received no special task instructions, thus forming the order (O) training group. Thus four different conditions of training were employed, EI, E, OI, and O.

A series of test anagrams was introduced immediately after the training series. All *Ss* received the same test which consisted of 10 five-letter anagrams each possessing several different words as solutions, one of which was an eating word. Each of the four training groups was divided into two subgroups for the test series. Half the *Ss* were instructed to look for eating words while half received no special task instructions. Thus eight different conditions were present in the test series. Three experimental variables were employed: training anagram, instructions prior to training, and instructions prior to the test series. They permitted a 2 × 2 × 2 factorial analysis of the number of eating or set solutions obtained on the multiple solution test anagrams.

Further details of the procedure and the instructions to *Ss* are similar to those in the experiment by Maltzman and Morrisett (3). The training and test anagrams were all obtained from the lists employed by Rees and Israel (4).

RESULTS AND DISCUSSION

The mean number of eating-set solutions obtained in the test series by the different groups is shown in Table 1. Bartlett's test indicated that the variances were homogeneous. A summary of the analysis of variance of the test data is given in Table 2. The analysis indicated that two of the

[1] This study was supported by a grant from the Committee on Research, University of California. A portion of this paper was presented at the 1954 meetings of the American Psychological Association in New York City.

[2] Now at the University of California, Riverside.

TABLE 1

MEAN EATING-SET SOLUTIONS TO TEST ANAGRAMS

Test Instructions	Training Instructions							
	Order Instructions		Order Only		Eating Instructions		Eating Only	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Eating instructions	6.75	1.36	6.42	2.09	7.17	1.33	7.58	1.34
Eating only	4.83	1.42	5.67	1.51	6.50	1.85	5.67	1.51

three main effects were significant: the class of anagrams presented during training and the nature of the instructions given prior to the test series. The nature of the instructions given prior to the training series did not significantly affect performance in the test series. None of the first- or second-order interactions were significant.

The significant *F* for training anagrams indicates that a set for eating-word solutions was established through the effects of training, a finding obtained in several other studies (**2, 4**). The *Ss* in the four subgroups receiving eating anagrams during the training series had a mean of 6.7 set solutions in the test series. The *Ss* in the four subgroups receiving practice in solving letter-order anagrams obtained a mean of 5.9 set solutions on the test series.

TABLE 2

ANALYSIS OF VARIANCE OF SET SOLUTIONS TO TEST ANAGRAMS

Source	*df*	Mean Square	*F*
Anagrams	1	15.84	5.91*
Training instructions	1	.01	
Test instructions	1	41.34	15.43**
Anagrams × Training instructions	1	1.26	
Anagrams × Test instructions	1	.01	
Training instructions × Test instructions	1	.01	
Anagrams × Training instructions × Test instructions	1	8.77	3.27
Within groups	88	2.68	

* $P = < .05$.
** $P = < .01$.

The task instructions given prior to the test series was the other variable producing a significant effect. The lack of significant interactions indicates that whether or not training anagrams and training task instructions were congruent with the test instructions, task instructions for eating solutions given immediately before the test produced a significant increase in this class of solutions.

The *Ss* in the four subgroups receiving test-task instructions had a mean of 7.0 set solutions. Fifty-four per cent of these *Ss* gave a set solution on the first test anagram. The *Ss* in the four subgroups without special task instructions before the test series obtained 5.7 mean set solutions. Twenty-five per cent of these *Ss* gave an eating solution on the first test anagram. A chi square corrected for continuity indicated that the relative frequency of set solutions obtained on the first test anagram by these different subgroups is significant (7.36, $P < .01$). These results indicate that the difference in set solutions obtained by the − EI and − E groups cannot be attributed to the acquisition of the set by the − EI subgroups during the course of the test series.

The immediacy of the effect of the task instructions is further indicated by a comparison of the different O-groups. Fifty per cent of the *Ss* in

the OI-EI and O-EI groups gave a set solution on the first anagram of the test series while 12.5% of the *Ss* in the OI-E and O-E groups gave a set solution ($\chi^2 = 6.21$, $P < .02$).

Although test-task instructions produced a significant increase in set solutions in the test series, training-task instructions did not have such an effect. A previous study by Maltzman and Morrisett (**3**) did obtain an increase in set solutions with task instructions given prior to the start of the training series. Rees and Israel (**4**) obtained comparable results in a similar experiment. However, the class of anagrams used in these studies—letter orders, and nature words —typically produce greater sets than eating words. This implies, of course, that the effectiveness of a given kind of task instruction will vary with the classes of anagrams employed. A determination of the generality of the effects of task instructions on set solutions would also require the systematic variation of the strengths of training and task-induced sets with different classes of anagrams.

Maltzman (**1**) has recently elaborated upon an interpretation of problem solving which may incorporate the phenomena traditionally described as mental set and determining tendency. Briefly, the theory assumes that mental sets established through training produce a change in the compound habit-family hierarchy by increasing the habit strength of a particular class of responses. Sets, or determining tendencies, established by task instructions produce a change in the compound habit-family hierarchy by increasing the reaction potential of a class of responses through arousal of their anticipatory goal response. Sets developed through training are based on the growth of habit strength, and therefore develop relatively slowly and have a degree of permanence. In contrast, sets induced by instructions are labile, are aroused rapidly, and need not persist. The results of the present experiment appear to be in accord with this theoretical interpretation.

Summary

The present experiment was designed to determine whether task instructions to look for a particular kind of solution to anagrams would induce a disposition for these solutions when previous training and instructions were incongruent with such a set. The results indicated that a set may be established by task instructions despite the incongruence of previous training and instructions. A theoretical analysis of sets induced by training and by instructions was outlined.

References

1. Maltzman, I. Thinking: from a behavioristic point of view. *Psychol. Rev.*, 1955, **66**, 275–286.
2. Maltzman, I., & Morrisett, L., Jr. Different strengths of set in the solution of anagrams. *J. exp. Psychol.*, 1952, **44**, 242–246.
3. Maltzman, I., & Morrisett, L., Jr. Effects of task instructions on solution of different classes of anagrams. *J. exp. Psychol.*, 1953, **45**, 351–354.
4. Rees, H., & Israel, H. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**, No. 6 (Whole No. 210).

(Received June 3, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

FURTHER GRADIENTS OF ERROR REINFORCEMENT FOLLOWING REPEATED REWARDED RESPONSES[1]

MELVIN H. MARX AND FELIX E. GOODSON

University of Missouri

In previously reported experimentation (**5, 7**) statistically significant spread-of-effect aftergradients were found in a Thorndikian (**11**) type of experimental situation with an apparatus designed to permit graded measurement of error strength and minimize systematic pre-experimental response sequences. Gradients were observed only after repeated rewarded responses. Bias in the response sequence, such as may be due to the operation of the guessing sequence (**2, 3, 8, 9**) or the probability-bias (**10**) factors, was controlled by using a control group of *S*s who also repeated, presumably by chance, the key response following which measurement of error strength was made.

A modified Thorndikian hypothesis was proposed as a consequence of these results. It was designed to take into account the fact that error strengthening occurred only following repetition of a rewarded response. It stressed the necessity of the repetition of the rewarded responses, as an important component in the stimulus complex present when the succeeding response is made.

It was the primary purpose of the present experiment to provide a further test of this particular hypothesis. This test was performed with a basically similar design but with a more naive type of *S* and an apparatus permitting the collection of group data.

[1] This research was supported in part by the United States Air Force under Contract No. AF33(038)-25631 monitored by the Perceptual and Motor Skill Research Laboratory of the Human Resources Research Center. Permission is granted for reproduction, translation, publication, use, and disposal by or for the United States Government. Part of the research described here was reported at the 1952 meetings of the American Psychological Association at Washington (abstracted in *Amer. Psychologist*, 1952, **7**, 274–275).

Method

Subjects

The *S*s were 385 fifth- and sixth-grade public-school students, used in groups of from 21 to 28. The experiment was performed in the classrooms of the several schools involved.

Apparatus

A small multiple-choice learning box was used. There were 27 rows of holes in the top, with 12 holes in each row. A second board was placed under this top. The second board was solid except for slits in some rows. These were the width of the holes above, and ran clear across the board under all 12 of the holes. A list of 27 three-letter stimulus words appeared at the left of a mimeographed record sheet covering the arrangement of holes in the top of the box. As *E* called out the stimulus word, *S* responded by inserting a stylus furnished for the purpose into one of the 12 dots on the paper to the right of the stimulus word. These dots were directly over the holes in the top board, so the stylus always pierced the paper, even on incorrect choices. When one of the slits in the second board was under the row of holes, the stylus went clear in to the hilt. The *S*s were told that this represented a correct choice. A hinged top covered the apparatus when it was not being used, and made it possible to change the record sheets without revealing the underlying arrangement to the *S*.

A major advantage of this apparatus was the elimination of a set central resting position to which *S* had to return following each response. This feature of the earlier apparatus (**5, 7**) had been a complicating factor in the interpretation of the gradients.

Experimental Procedure

The *S*s were told that the test was one of their ability to learn and remember the correct hole

for as many of the words as possible. A number of items in the instructions were used to stimulate motivation and disguise the purpose of the experiment.[2]

Experimental *Ss* were rewarded according to one of three prearranged conditions on the first two trials. They received rewards on either (*a*) the 14th response, (*b*) the 5th and 23rd responses, or (*c*) the 5th, 14th, and 23rd responses. The positions of reward were determined by the slits in the second board of the apparatus. Different sets of boxes had slits under different rows. One set had slits under Row 14, another under 5 and 23, and a third under 5, 14, and 23. There were no slits in the boards for control *Ss*, who therefore received no reward on either of these two trials. These conditions are designated as 1-reward ($N = 67$), 2-reward ($N = 57$), 3-reward ($N = 127$), and 0-reward ($N = 138$) groups, respectively.

After the introductory instructions, *Ss* were shown a sample learning box in order to familiarize them with details of construction. This box had a random arrangement of correct holes. Each *S* was then given a small box containing only two holes through one of which the stylus passed up to the hilt and into one of which the stylus would go only ¼ in. Each *S* punched these holes with his stylus in order to get the feel of the right and wrong signals.

Trial 1.—The *Ss* were now given their full-sized boxes, together with a 5 × 8-in. blank card. They were told to punch once and only once in each row, and to cover previous punches with the card as they proceeded down the page. The *Ss* were asked not to talk to each other at any time during the experiment. They were also told not to look beneath the sheet of paper in the box at any time. Proctors assisted *E* to insure that these cautions were observed.

The *Ss* were assured that this was a very difficult task and that they would probably not get many right, if any, but that they should keep trying. They were advised that if they guessed randomly rather than by some systematic method they would probably do better. The words were read off by *E* at the rate of one every 8 sec. A trial was completed when *Ss* had punched once in each of the 27 rows as the 27 stimulus words were read off.

Trial 2.—The sheets used on Trial 1 were removed by *E*'s assistants and new sheets were placed in the boxes. This was done with the tops of the boxes down so that *S* could not see into the boxes. The *Ss* were told that on this trial they were to do two things: (*a*) repeat all those dots which they had gotten right on the first trial, and (*b*) try to get more new ones right. Instructions about talking, looking into boxes, etc. were repeated on this and the last two trials.

Trial 3.—At the end of Trial 2 the boxes were interchanged among *Ss*. 0-reward and 3-reward *Ss* exchanged boxes, as did the 1- and 2-reward groups. These two additional trials with the change in correct responses were introduced to try to prevent discouragement among the 0-reward *Ss*. Trials 3 and 4 did not provide a good test of the hypothesis, because of interference from earlier trials and possible response selection factors resulting from continuation past the second trial. Nevertheless, the data were analyzed for whatever additional information they might give. The same stimulus terms were used on Trials 3 and 4, which did not tend to minimize interference.

Trial 4.—Trial 4 was exactly like Trial 2. The *Ss* were told to repeat all those responses that they had gotten correct on Trial 3 and to try to make further correct responses.

Treatment of Data

The columns of dots punched by *Ss* on the record sheets were arbitrarily numbered from 1 to 12 from left to right. Each *S*'s 27 responses on Trials 1 through 4 were recorded in terms of these numbers. Then the difference between the number of each response on Trial 1 and the corresponding response number on Trial 2 was computed. These difference scores, which ranged from 0 to 11, provided a quantitative expression of the strength of the tendency to repeat a given response on successive trials. A low score thus indicated a strong tendency to repeat a response. This method is essentially the same as the difference-score measure previously used (**5, 7**; cf. also **6**).

Within the experimental groups *Ss* were assigned to subgroups on the basis of their difference scores at the rewarded positions. The subgroups thus formed were: (*a*) *repeaters*—those *Ss* having a difference score of 0 at one of the rewarded positions; (*b*) *near-repeaters*—those *Ss* having a difference score of 1 at one of the rewarded positions; and (*c*) *nonrepeaters*—those *Ss* having a score of 2 or greater at one of the rewarded positions. This breakdown was planned on the basis of previous results (**5, 7**) which indicated the similarity of exact repeaters and near-repeaters, as here defined.

Since some experimental groups received more than one reward per trial it was, of course,

[2] The verbatim instructions for the experiment, along with tabulations and analyses of data summarized in the text, have been deposited with the American Documentation Institute. Order Document No. 4815, remitting $1.75 for 35-mm. microfilm or $2.50 for 6 by 8-in. photocopies.

possible for an *S* to have more than one score of 0, 1, or 2 to 11 at key positions. Only the first such score was included in the appropriate analysis. No *S* appears more than once in any of the groups to which a statistical test has been applied. However, it is possible for a portion of an *S*'s responses to appear in more than one analysis. That is to say, since no direct comparisons have been made among repeaters, near-repeaters, and nonrepeaters, an *S* having a 0 score at one rewarded position and a 1 or a 2–11 score at another may appear in two or three of these experimental subgroups.

In order to obtain homogeneity of variance, it was necessary to transform the scores by taking the common log of $(X + .5)$.

Portions of some protocols were discarded in advance of analysis. The criteria for discarding data were: (*a*) failure to make a response in one of the critical positions, that is, in a rewarded-response position or a nearby error position; (*b*) making more than one response in any of the critical positions; and (*c*) using obvious guessing systems such as are indicated by Dots 1, 2, 3, 4, or 4, 6, 8, 6, 4. This last precaution was especially important in the present study because of the impracticability of emphasizing, with grade-school *S*s, the customary instruction to "guess at random" (cf. **6**). The present instructions did urge *S*s to use different numbers in guessing.

Results

Repeaters

The method of obtaining experimental repeaters has been described. Control repeaters were obtained from the 0-reward group by taking the difference scores for four steps after the *first* repeated response between the 5th and 23rd positions, in case there was more than one such repetition. In rare cases where no repetition occurred, the *S*'s data could not be used in this analysis. Earlier results (**5, 7**) indicate that the serial position of the repeated response is not crucial in determining whether a gradient will occur. There were 102 experimental and 112 control repeaters.

The experimental and control aftergradients for Trials 1 and 2 are shown in Fig. 1. It will be noted that a regular and definite aftergradient occurred in the experimental group and that only a slight aftergradient occurred in the control group. No experimental foregradient was evident.

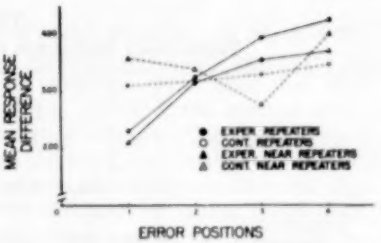

FIG. 1. Response-difference aftergradients for experimental and control repeaters and near-repeaters, Trials 1 and 2.

Transformed mean scores and variances for the four aftergradient positions are given in Table 1. For purpose of comparison this table also contains the data for near-repeaters and nonrepeaters. The trend test described by Linquist (**4,** Case 2) was applied to the transformed-score gradients. Only the experimental aftergradient departed significantly from a horizontal straight line ($F = 8.92$, $P = .001$ for 3 and 303 *df*).

A direct comparison between experimental and control aftergradients was made following Lindquist's suggested method (**4,** Case 11). This analysis of variance indicated a statistically significant difference in slope ($F = 3.349$, $P = .01$ for 3 and 636 *df*). In addition, the experimental mean difference score was significantly lower than the control group mean in the first aftererror position ($F = 6.46$, $P = .05$ for 1 and 212 *df*).

The repeaters' data were also analyzed for exact error repetitions, which are the orthodox measures used in the earlier spread-of-effect studies. Cochran's (**1**) chi-square test for correlated variables was first applied

TABLE 1

TRANSFORMED MEANS FOR AFTERGRADIENTS, TRIALS 1 AND 2

Subgroup	N	Error Positions							
		+1		+2		+3		+4	
		Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Experimental Repeaters	102	.2845	.1417	.4226	.1441	.5255	.1269	.5231	.1510
Control Repeaters	112	.4311	.1385	.4385	.1355	.5447	.1389	.4666	.1598
Experimental Near-repeaters	52	.2612	.1357	.4578	.1118	.4572	.1204	.4640	.1241
Control Near-repeaters	51	.4488	.1593	.4146	.2069	.3646	.1400	.4816	.1109
Experimental Nonrepeaters	129	.3517	.1606	.4627	.1169	.4811	.1461	.5057	.1213
Control Nonrepeaters	126	.3961	.1393	.5064	.1241	.5254	.1418	.4600	.1511

independently to the fore- and aftergradients. The results of this test confirmed those obtained by trend tests based upon the difference scores. Also, there were significantly more error repetitions in the first aftererror position for the experimental group.

A further measure of the tendency to repeat errors was obtained by correlating the responses for each error position on Trial 1 with the responses at that position on Trial 2. Table 2 presents the correlations and tests of the reliability of the differences between experimental and control r's.

TABLE 2

CORRELATIONS BETWEEN ERROR RESPONSES OF REPEATERS, TRIALS 1 AND 2

Error Position	Experimental		Control		Difference	
	r	z'	r	z'	z'	t
−4	.148	.149				
−3	.078	.078				
−2	.126	.127				
−1	.057	.057				
+1	.380*	.400	−.054	−.054	.454	3.27*
+2	.024	.024	.135	.136	.112	0.81
+3	−.128	−.129	.039	.039	.168	1.21
+4	−.099	−.099	.087	.087	.012	0.09

* $P = .01$.

The only significant difference in correlations between the two groups was at the first aftererror position, where the experimental r of .38 was itself significantly above zero at the .01 level. These results corroborate those obtained by the difference-score method and indicate that the preceding reward is a necessary condition for the repetition of errors. The correlation measure is especially important as a supplement to the difference-score measure because it is relatively insensitive to departures from normality, and is therefore fairly free of any distributional artifacts arising as a result of reward.

In order to eliminate any group guessing sequences as alternative explanations of the major results a partial correlation technique was also applied. When the correlations between the rewarded response position and the first aftererror position on each trial were partialled out, the correlation between the two first aftererror positions remained significant. The r was only slightly reduced, to .34, still significant at the .01 level. This means that the pos-

sible concurrence on both trials of a relationship between the location of the rewarded response and the error immediately following cannot be used as an explanation of the strength of this error.

A check was also made of the possibility that some side effect of reward operating on the immediately following afterresponses on both trials was responsible for the first aftererror strengthening. The most important possibility here seemed to be that following a reward *S* tended to restrict the next response to the same general area. The partial correlation technique takes care of this possibility in part, but a more direct test was also made. The experimental and control aftergradients were broken down according to the difference between the repeated response and the first aftererror on Trial 1. Examination of these data indicated that (*a*) experimental and control *S*s showed about the same tendency to stay close to the repeated response, and (*b*) the first aftererror repetition tendency was greater in all experimental classifications except the last one. These facts suggest that artificial response restriction following reward was not an important gradient-producing factor.

The experimental data in terms of difference-score gradients around repeated rewarded responses were analyzed separately for each of the three serial positions, 5, 14, and 23. All repeated rewarded responses at each position were used for this analysis. Each of the marked aftergradients was significant at beyond the .01 level, as shown in Table 3. These results indicate that the significantly positive results based upon first occurrence of the repeated rewarded response and compared with repeated nonrewarded controls were not a function of serial position. The foregradients were irregular and did not justify statistical test.

Near-Repeaters

Gradients of error strength around near-repetitions (that is, intertrial variations in the correct response of only one unit) were also analyzed on the assumption that most such cases probably represented very close approximations to successful repetition. A number of such near-misses would be expected in view of the lack of numerical or other definition on the response scale. Experimental near-repeaters were selected in the same manner as experimental repeaters. Difference scores were recorded for four steps before and after the first difference score of 1 occurring in a rewarded-response position. Control near-repeaters were obtained by first locating all control *S*s with difference scores of 1 at Positions 5, 14, or 23—the equivalents of the rewarded

TABLE 3

AFTERGRADIENTS OF ALL REPEATERS AT EACH REWARDED POSITION, TRIALS 1 AND 2

Reinforced Serial Position	*N*	Mean Difference Score by Error Position				Trend Test*	
		+1	+2	+3	+4	*F*	*P*
5	42	2.57	2.69	4.61	3.83	15.69	.01
14	44	2.07	3.52	3.89	3.45	5.47	.01
23	42	2.23	2.95	3.35	4.35	7.40	.01

* Based on transformed scores.

positions in the experimental groups. The first such difference score was used, except for those *Ss* who happened to have three 1 scores, in which cases the selection was made at random. There were 52 experimental and 51 control near-repeaters thus obtained.

The aftergradients for the near-repeaters are shown in Fig. 1. Again the experimental aftergradient was more marked than the control. Trend tests were made of all gradients, just as they were for repeaters. Again, only the experimental aftergradient was significantly different from a straight horizontal line ($F = 6.15$, $P = .01$ for 3 and 153 *df*).

The over-all analyses of variance on transformed scores yielded a significant difference between the experimental and control aftergradients ($F = 3.42$, $P = .02$ for 3 and 303 *df*). The experimental-control difference between the mean transformed difference scores at the first aftererror position was also significant ($F = 6.01$, $P = .02$ for 1 and 101 *df*).

Nonrepeaters

These *Ss* were selected in the same manner as experimental and control repeaters, i.e., the first score of 2 through 11 that occurred in response Position 5, 14, or 23 was used to represent each *S*. Raw-score gradients for first to fourth errors were as follows: 2.68, 3.19, 3.60, and 3.57 for the experimental *Ss*; 3.02, 3.64, 3.87, and 3.45 for the control *Ss*.

Transformed aftergradient data for experimental and control nonrepeaters are given in Table 1. Although relatively weak compared with experimental repeaters' and near-repeaters' gradients, both of these were nonetheless statistically significant (experimental $F = 5.58$, $P = .01$ for 3 and 384 *df*; control $F = 3.32$, $P = .05$ for 3 and 375 *df*). A direct comparison of the two gradients showed the experimental to be stronger at the .06 level of significance ($F = 2.54$, for 3 and 756 *df*; 2.62 required for .05 for 3 and 500 *df*). The intertrial r for raw first aftererror responses was .15 for the experimental *Ss*, just short of significance at the .05 level (r of .17 required for N of 125). The r for the control *Ss* was .01.

Trials 3 and 4

The analysis of the data provided by Trials 3 and 4 was conducted in exactly the same manner as that of the data from Trials 1 and 2. The major analyses are here briefly reported.

Trend tests were made of the repeaters' gradients from 115 experimental and 67 control *Ss*. Only the experimental aftergradient departed significantly from a straight line ($F = 9.66$, $P = .01$ for 3 and 342 *df*).

Direct comparisons of the slopes and means of the experimental and control groups revealed no significant differences. Nor was the mean difference in the difference scores at the first aftererror position significant ($F = 2.57$; need $F = 3.89$ for $P = .05$).

The correlation between responses on Trials 3 and 4 at the first aftererror position for the experimental *Ss* was .27 ($P = .01$), and for the control was not significant ($r = .16$). The difference between these correlations was not significant.

The partial correlation technique was applied to the experimental group correlation, partialling out the correlations between the location of the rewarded response and of first aftererror responses. The resulting intertrial correlation was .26, still significant at the .01 level.

Near-repeater gradients were based upon 57 experimental and 35 control *Ss*. None of the gradients was significantly different from a straight line. The experimental aftergradient, however, was significantly stronger than the control in terms of level ($F = 6.33$, $P = .02$ for 1 and 270 *df*). The difference between the difference-score means at the first aftererror position also significantly favored the experimental group ($F = 12.77$, $P = .01$ for 1 and 90 *df*).

Nonrepeater gradients were based upon 108 experimental and 76 control *Ss*. Only the experimental aftergradient was statistically significant ($F = 3.62$, $P = .02$ for 3 and 321 *df*).

It did not differ significantly, however, from the control gradient.

Discussion

The present results offer confirmation of those previously obtained (**5, 7**) with a different type of *S* and apparatus. Trials 1 and 2, which constituted a clear test, gave the strongest confirmation. It is especially important to note that direct comparisons of the experimental and control data yielded statistically significant differences. Less clear-cut support was given by Trials 3 and 4, where there were contaminating factors.

The finding of significant aftergradients following *nonrepeated* key responses requires comment. This result was unexpected on the basis of the previous results. The fact that experimental aftergradients following nonrepeated rewarded responses occurred both on Trials 1–2 and 3–4 offers support for Thorndike's original "spread" hypothesis (**11**). This is particularly true since the experimental aftergradient on Trials 1–2 was found to be stronger than the control at the .06 level. If verified, this result would eliminate the need for the modified Thorndikian hypothesis previously advanced.

The fact that even the control aftergradient following nonrepeated key responses was significant on Trials 1–2, suggested that perhaps there was something unusual about Positions 5, 14, and 23, since this result would not be predicted on the basis of the guessing-sequence or similar response-bias hypothesis which involves repeated responses as anchors. A special check of the control nonrepeater aftergradient following intermediate response Positions 9 and 18 indirectly supported this suggestion, since no tendency towards a gradient was observed (raw mean difference scores of 4.29, 3.61, 3.70, and 3.44 for first through fourth after position). A second special check was made on control *S*s who had repeated errors in the key positions 5, 14, and 23. Only 44 *S*s were found who had repeated at one of these positions, and their aftergradient was essentially similar to that observed for the larger control subgroup based upon all response positions (3.00, 2.89, 3.50, and 3.45). This gradient was not significant ($F < 1.00$), nor was the experimental repeater aftergradient significantly stronger than it ($F = 1.32$, $P > .05$ for 3 and 432 *df*). However' it should also be noted that the control near-repeaters, based upon response positions 5, 14, and 23, failed to provide a significant aftergradient. These considerations suggest that this single significant control aftergradient found for nonrepeaters may most plausibly be regarded as a sampling variation.

This conclusion is fortified by the failure to find any intertrial correlation ($r = .01$) for the first control aftererror position. But such considerations leave unresolved the problem of the two significant experimental nonrepeater aftergradients, which are equally embarrassing to the modified Thorndikian (**5, 7**) and the guessing-sequence (**3**) hypotheses since each depends upon repetitions of key responses. The present results at least suggest that this problem be carefully considered in subsequent research and that the necessity of repetition of key response not be arbitrarily assumed without checking the data.

With the qualifications imposed by the above comments, the results are interpreted as offering support for the modified Thorndikian spread-of-effect hypothesis previously advanced. The necessity for reward is indicated by the failure to find statistically significant control gradients following repeated nonrewarded responses. This control and the use of the partial correlation technique in the present study eliminate the guessing-sequence and the probability-bias hypotheses as possible explanations of the positive results.

The finding of a significant correlation for only the first aftererror raises a question as to the extent of the spread of reward. Whether or not the aftergradient extends beyond the first step is a question which must be decided by further experimentation.

Summary

This study involved the analysis of data provided by 385 fifth- and sixth-grade public-school children who were given four trials on an orthodox Thorndikian type of multiple-choice learning task with prearranged conditions of reward. In general it was found that the errors made immediately following a repeated rewarded

response tended to be significantly stronger than comparable errors made immediately following a repeated but nonrewarded response. The guessing-sequence factor and similar nonrandom biasing of the response sequences were controlled by use of repeated nonrewarded responses as well as by application of a partial-correlation technique. Some tendencies toward after-gradients were also found following nonrepeated responses, however, constituting support for the original Thorndikian spread-of-effect hypothesis. With this exception, the results are interpreted as confirmation of similar ones previously obtained with a different apparatus and a less naive type of *S*, and as offering support for the modified Thorndikian spread-of-effect hypothesis previously advanced.

References

1. Cochran, W. G. The comparison of percentages in matched samples. *Biometrika*, 1950, **41**, 256–266.
2. Fagan, C. A., & North, A. J. A verification of the guessing sequence hypothesis about spread of effect. *J. exp. Psychol.*, 1951, **41**, 349–351.
3. Jenkins, W. O., & Sheffield, F. C. Rehearsal and guessing habits as sources of the "spread of effect." *J. exp. Psychol.*, 1946, **36**, 316–330.
4. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
5. Marx, M. H. Error-reinforcement gradients with bias in the response series controlled. *Amer. Psychologist*, 1951, **6**, 293. (Abstract)
6. Marx, M. H. Spread of effect: a critical review. *Genet. psychol. Monogr.*, 1956, **53**, 119–186.
7. Marx, M. H. Gradients of error reinforcement in a serial perceptual-motor task. *Psychol. Monogr.*, in press.
8. Sheffield, F. D. "Spread of effect" without reward or learning. *J. exp. Psychol.*, 1949, **39**, 575–579.
9. Sheffield, F. D., & Jenkins, W. O. Level of repetition in the "spread of effect." *J. exp. Psychol.*, 1942, **44**, 101–107.
10. Smith, M. H. Spread of effect is the spurious result of non-random response tendencies. *J. exp. Psychol.*, 1949, **39**, 355–368.
11. Thorndike, E. L. An experimental study of rewards. *Teach. Coll. Contr. Educ.*, 1933, No. 580.

(Received April 18, 1955)

Journal of Experimental Psychology
Vol. 51, No. 6, 1956

SAMPLED-DATA TRACKING: SAMPLING OF THE OPERATOR'S OUTPUT [1]

CORWIN A. BENNETT

Human Factors Group, International Business Machines Corporation

In the usual display-control situation under normal operating conditions, the operator closes a control loop by tracking a target. Thus, in a compensatory tracking task, when the operator sees a discrepancy between a movable target and fixed cross hairs, he displaces his control in order to bring the target back onto the cross hairs. The operator's output or response is then continuously sensed and utilized by analog equipment. However, when such an analog device is replaced by a digital computer obtaining solutions at finite time intervals, the operator is placed in a unique situation. The digital computer can accept the output of the operator as an input only at finite time intervals. Accordingly, the computer will present solutions as an information input or display to the operator only at such intervals. Thus, the operator, acting near-continuously on an analog control device which is sampled at discrete time intervals by the digital computing equipment, sees the results of his activities as discrete events on the display, smoothed only by whatever inertia may be present in the control mechanism.[2] It is pertinent to ask what effects such a digitalization will have on the operator's tracking achievements. Particularly, one might anticipate a degradation of tracking performance under such sampled-data conditions.

Previous experiments (**1, 2, 4**), in which sampling or intermittency was present in the control loop, included sampling of the input to the operator rather than his output. From the operator's standpoint, however, there is a difference between sampling his input and sampling his output. In sampling his input, the occurrence and result of sampling are immediately apparent. In sampling his output, sampling effects are noticeable only after the operator's response is fed back to him through the computer and display.[3] Figure 1 shows the general nature of the control loop for the two types of sampling conditions.[4]

The primary variable entailed by control-loop sampling is the sampling rate. Battig *et al.* (**1**) studied rates between .4 cps and 16.1 cps. An

[1] The writer would like to acknowledge the contributions to various phases of this work of M. W. Olson, M. M. Dickinson, R. M. Lewis, and J. S. Renfro of IBM's Airborne Computer Laboratory; W. J. Merrill and P. M. Seever of the Scientific Computation Laboratory, and W. L. Jenkins of Lehigh University.

[2] This smoothing is such that, without inertia, a moving target appears as a succession of dots on the display. Thus, the target is stationary except at sampling time when it jumps to another position. With sufficient inertia the moving target appears as a continuous trace. Further experimentation is being conducted on the effects of amount of inertia on sampled-data tracking performance.

[3] These delays due to sampling may vary from one to two sampling periods in length. Thus, at a 2-cps rate, the delay between operator's response and the displayed effect is at least ½ sec. and no more than 1 sec. In addition to sampling delays, there are exponential delays in the present experimentation due to system inertia such that after .25 sec., approximately two-thirds of the total target displacement will have taken place.

[4] Sampling might well take place at *both* the operator's input (display) and output (control). Studies are currently being performed on such a control system.

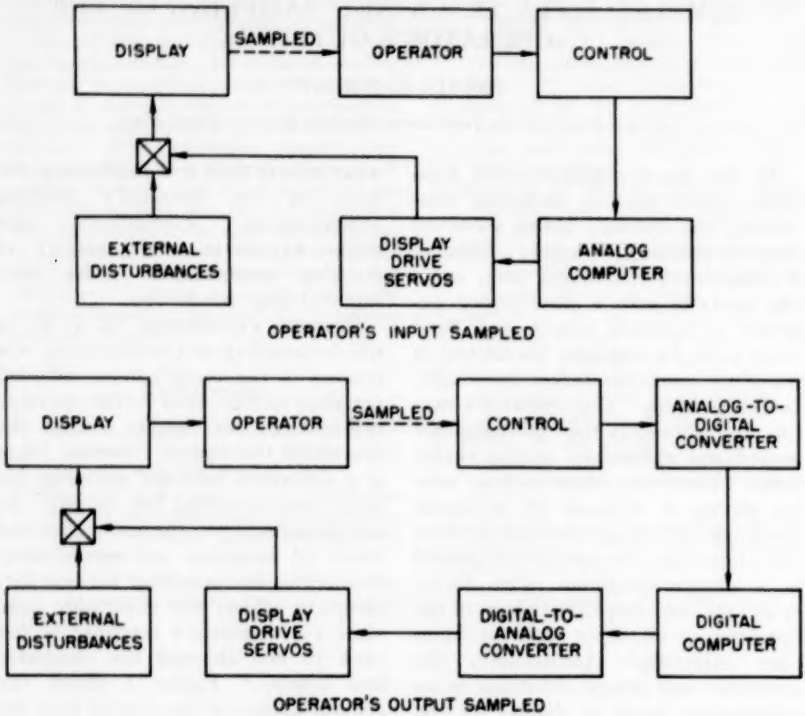

FIG. 1. Control loop block diagrams: Sampling of operator's input and output.

optimal rate between 6 and 16 cps was found with large decrements in time on target at lower rates. This is contrary to the previously implied expectation that with an infinite sampling rate (continuous), performance would be optimum, and, as sampling rate is reduced, performance would be degraded. This experiment was repeated (**2**) with control of previously uncontrolled variables. No optimum rate was found, but a linear increase in performance for rates from 2 to 16 cps.

Katz and Spragg (**4**) studied mean tracking error as a function of intermittency as related to operator's input sampling. They describe their findings as showing a positive logarithmic relationship between tracking accuracy and illumination frequency. In another experiment described in the same report, a 4-cps rate and a continuous condition were investigated. The continuous condition proved to be superior.

In addition to the effects of sampling in the control loop, one might wish to determine the effects of introducing constant delay periods between the operator's response and effects of the response as presented on the display. Such delays have have been called "transmission-type delays" and were demonstrated by Warrick (**8**) to result in degraded

tracking performance in one type of tracking situation. Rockway (**6**) has reviewed studies of the effects of this and other types of delays on tracking performance.

A further problem common to all control devices is that of "control/display ratio," or, "sensitivity" of the control. Control/display ratio may be defined as the magnitude of displacement (or one of its derivatives) on the display, produced by a given control displacement.[5] The sensitivity might be expected to be of especial importance in the sampled-data situation.

One further variable was introduced into the experimentation, that of tracking task. One type of tracking task arises when the target is instantaneously displaced (by an external disturbance) from the cross hairs—"discretely displaced target." The task then is to return the target as quickly as possible to the cross hairs. Another tracking task arises when the target is slowly but continually displaced at a given rate from the cross hairs—"constantly displaced target." The operator's task is then to set a rate into the system equal in magnitude and opposite in direction to the disturbance, and thus maintain the target on the cross hairs.

It should be pointed out that, with the present instrumentation, a given displacement of the operator's control moves the target at a given rate. Therefore, the first task requires the operator to move the discretely displaced target at some rate, releasing the deadman switch at the appropriate time to stop the target on the cross hairs. The second task requires the operator to try to produce a rate equal and opposite to the disturbing rate. By the time such a rate is discovered, however, the target has been displaced from the cross hairs. Thus, the operator must simultaneously try to remove both rate and position errors.

In summary, the present problems in regard to sampled-data tracking are the determination of: (*a*) Tracking performance as a function of sampling rate and hand control sensitivity for discretely displaced targets; (*b*) tracking performance as a function of sampling rate for delay and no-delay conditions for discretely displaced targets; and (*c*) tracking performance as a function of sampling rate for constantly displaced targets. Three experiments were carried out to determine these three functions.

Method

Apparatus.—The simulation of the digital control loop in which *S*'s output is sampled is shown in Fig. 2. The digital computer was simulated by an analog computer (a modified "integral-differential analyzer") and relays. In this simulation sampling "on time," during which signals were "accepted," was approximately 15 msec. Sampling "off time" was the reciprocal of sampling rate (in cps) in seconds minus "on time."

The task was a non-aided compensatory tracking situation. The hand control used was a spring-centered rate-controlled joy-stick (8-in. radius), equipped with a dead-man switch. Forces required for the displacement of the control were .5 lb. initially and 2.5 lb. for total displacement.

The display was a $\frac{1}{16}$-in. spot presented on a DuMont 5-in. laboratory cathode-ray oscilloscope (Type 804-A) with a P-1 phosphor. The scope was provided with cross hairs and three concentric markers—$\frac{1}{2}$, 1, and 2 in. in radius. The equipment was located in a semi-darkened room. The *S*s were allowed to set

[5] Some writers prefer "display/control ratio" or "magnitude of hand-control displacement required to effect a given rate of movement on the display." The author follows Chapanis *et al.* (**3**, pp. 325, 328), and more recently, Rockway (**6**). Rockway summarizes investigation of control/display ratio, but unfortunately these investigations have no necessary bearing on the present situation since they have been concerned with *position* (displacement) control rather than *rate* control.

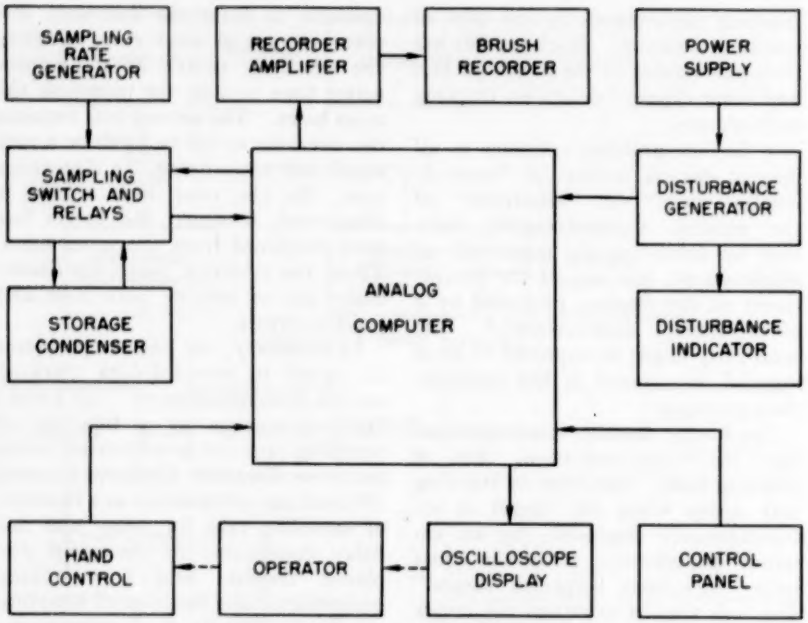

FIG. 2. Simulated digital control system.

their viewing distance. These distances varied between 12 and 16 in.

Sensitivity changes were accomplished by potentiometer adjustments in the analog computer. Since the external disturbances entered the loop subsequent to these potentiometers, changes in gain sensitivity had no effect on what *S* saw in the way of disturbances.

Target position was plotted as a function of time on a Brush Recorder. Measurements of the criteria (recovery time and 50% limits) were made manually from these records.

Subjects.—The *Ss* were an incidental sample of engineers and technicians. The same seven *Ss* were used in Exp. 1 and 2. The seven *Ss* used in Exp. 3 had previously taken part in either the first two studies or in a similar study conducted previously. Those *Ss* unacquainted with the purposes of the experimentation were familiarized with them at the beginning of the study.

Experiment 1.—Experiment 1 was designed to determine the relationship between sampling rate, hand-control sensitivity, and tracking performance for discretely displaced targets.

The choice of sampling rates to be studied was based on the results of a pilot study. In that study, a 2-cps rate seemed a limiting case beyond which tracking was unreasonably poor. Other sampling rates, chosen in the region at which satisfactory tracking was expected, were 3, 4, and 6 cps. A continuous (nonsampled, analog) condition was included as a reference situation.

Experience with sampled-data tracking prior to the experimentation indicated that as sampling rate decreased, the optimum hand-control sensitivity decreased; that is to say, for lower sampling rates, it was desirable to have a given control movement effect less display movement. For any given sampling rate, it was expected that an optimum sensitivity would exist, and either a higher or a lower sensitivity would result in poorer tracking. Three sensitivities were chosen so that an apparently optimum, a high but reasonable, and a low but reasonable sensitivity for each sampling rate would be used.[6] Specific sensitivities used for each sampling rate are shown in Table 1.

[6] Originally it was intended that a wide enough range of sensitivities to include the optimal sensitivity for each sampling rate (studied) would be included in a complete factorial design. For reasons of economy of time this was not done.

TABLE 1

SENSITIVITIES AND SAMPLING RATES USED IN EXP. 1

Sampling Rate (cps)	Sensitivity, Inches/Second/20° Hand-Control Deflection		
	Low Rate	Optimal	High Rate
2	1.0	1.5	2.5
3	1.0	1.5	2.5
4	1.0	1.5	3.0
6	1.0	1.5	4.0
Continuous	1.5	3.0	4.0

"Recovery time" was used as the criterion of tracking performance. Recovery time is defined as the time from the introduction of the disturbance until *S* brings the target within the ¼-in. circle and maintains it there.

Five series of eight disturbances each were used as the tracking tasks. A disturbance would instantaneously drive the target to the 1-in. circle at one of the cardinal headings, or to the 2-in. circle at the cardinal or intercardinal headings. Each series was made up of four disturbances of each magnitude, with heading and magnitude randomized.

Tracking was performed on 12 successive work days, with 7 complete replications of the 15 conditions for each *S*. The first three replications served as training, during which the conditions were presented in sequence to ease the transition from one condition to the next. On the last four replications, conditions were presented in random order to eliminate any systematic effects of change from condition to condition. Data from only these last four replications were used in the analysis of results.

On any given day, *S* operated under about 9 of the 15 different conditions of sampling rate and sensitivity. The *Ss* were instructed to bring the target within the ¼-in. circle as quickly as possible. Before any given condition, *S* was given a 30-sec. practice session, during which he was allowed to move the target at will and thus get the "feel" of the control condition (as an attempt at eliminating order-of-conditions effects). Subsequent to the practice period, a ready signal was given and the trial started. The disturbances were presented at 12-sec. intervals during the training sessions, and at 10-sec. intervals thereafter.

Experiment 2.—The second experiment immediately followed the first. The same *Ss*, disturbances, and criterion of tracking performance were used. This study was designed to determine the effects of delay on tracking performance as a function of sampling rate.

In this experiment, a constant delay period equal to the sampling interval (e.g., .33 sec. for 3-cps sampling) was introduced between *Ss* response and indications of the response on the display. Based on the results of Exp. 1, the following sampling rates were chosen: 3, 4, 5, 6, and 8 cps. Each of these sampling rates was studied with and without the additional delay period. Indications were that, as optimum sensitivity decreases with sampling rate, it also decreases with delay. One sensitivity was chosen for each sampling rate to be optimum for the delay condition (and, thus, less than optimum for the no-delay condition). Thus, any decrement in performance using the delay conditions could not be attributed to sensitivity. These sensitivities were: 1, 1, 1.5, 2, and 2.5 in. per sec. per 20° of hand-control displacement.

The training period again consisted of three replications, but for the delay conditions only, since *Ss* already had experience under the no-delay conditions. In the final four replications, *Ss* performed under the various delay, no-delay conditions in random order. These last four replications were used in the analysis of results. The seven replications required seven working days.

Experiment 3.—The third experiment was designed to determine the relationship between sampling rate and tracking performance for targets which are slowly but constantly displaced from the cross hairs.

Preliminary experience with this tracking task indicated two things in contrast to the previous tracking situation: (*a*) little, if any, learning would take place; and (*b*) lower sampling rates and sensitivities would yield tracking performance on a par with the continuous or analog condition.

Five sampling conditions were used: 1, 2, 3, and 5 cps, and a continuous condition. The sensitivity used for all conditions was .1 in. per sec. per 20° of hand-control displacement.

The "50% limits" was used as the criterion of performance. The "50% limits" is defined as the distance in both horizontal and vertical dimensions on the display within which *S* was able to maintain the target 50% of the time. With rate tracking of targets displaced at a constant rate, the error-response record was more or less continuously varying about the calibrated zero-error line rather than being discontinuous. Recovery time was thus not as meaningful a measure for such tracking.

A new set of 20 disturbances was made up, each .02 in. per sec. in magnitude. The disturbances were in the eight cardinal and intercardinal directions with equal frequencies. The series was constructed so that each pair of disturbances was separated by an interval of no

disturbance. Each disturbance changed direction 45° from the previous disturbance. Two sessions, on successive days, were given the *Ss*. Each session consisted of performance under each of the five sampling conditions (again, in random order) for 2 min., during which time four 15-sec. disturbances separated by four 15-sec. intervals were presented.

Results

Analysis of variance.—An analysis of variance was carried out on the data from each of the three experiments. Tables 2, 3, and 4 show the results of these analyses for main effects and all significant interactions. Estimates of the component of variance attributable to each factor (Column 6) and the percentage of the sum of the variance components attributable to each factor (Column 7) were computed for each experiment.[7]

In Exp. 1 (Table 2) seven of the factors affect tracking as indicated

[7] Discussion and computation of components of variance may be found in Mood (**5**, pp. 342 ff.). The percentage of the sum of components of variance (Column 7) is simply the individual component divided by the sum of Column 6, quotient times 100.

TABLE 2

Summary of Analysis of Variance: Exp. 1

Source	*df*	Mean Square	*V*‡†	*F*	Estimated Component of Variance	% Sum of Components of Variance
Subjects (A)	6	367.00	5, 6, 7, 9‡	19.47**	0.732	37
Sampling rate (B)	4	705.00	5, 8, 9‡	46.50**	1.026	52
Sensitivity (C)	2	30.40	6, 8, 9‡	2.04		
Sessions (D)	3	7.69	7	0.81		
A × B	24	6.44	9	1.96**	0.033	2
A × C	12	6.15	9	1.88**	0.018	1
A × D	18	9.54	9	2.91**	0.052	3
B × C	8	12.00	9	3.66**	0.039	2
A × B × C × D	144	3.28	10	1.21*	0.070	3
Pooled residual	3138	2.72				
Total	3359				1.970	100

* Significant at .05 level.
** Significant at .01 level.
† Denominator in *F* test.
‡ For tests of effects involving several significant components, see Walker and Lev (**7**, pp. 273–3).

TABLE 3

Summary of Analysis of Variance: Exp. 2

Source	*df*	Mean Square	*V*†	*F*	Estimated Component of Variance	% Sum of Components of Variance
Subjects (A)	6	164.73	6	70.70**	0.508	42
Sampling rates (B)	4	155.17	5	14.08*	0.322	27
Delay-no delay (C)	1	287.93	5	26.10**	0.247	21
Sessions (D)	3	8.33	6	3.58*	0.011	1
B × C	4	11.03	6	4.73**	0.039	3
A × B × C × D	72	2.33	7	1.35*	0.076	6
Pooled residual	2149	1.72				
Total	2239				1.203	100

* Significant at .05 level.
** Significant at .01 level.
† Denominator in *F* test.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE: EXP. 3

Source	*df*	Mean Square	*Vs*†	*F*	Estimated Component of Variance	% Sum of Components of Variance
Subjects (A)	6	4.22	4	3.12		
Sampling rates (B)	4	26.35	5	51.67**	1.84	92
Sessions (C)	1	2.71	4	2.01		
A × C	6	1.35	5	2.65*	0.17	8
Pooled residual	52	0 51				
Total	69				2.01	100

* Significant at .05 level.
** Significant at .01 level.
† Denominator in *F* test.

by the *F* tests. Two of these factors, *Ss* and Sampling, accounted for nearly 90% of the explained variance. The remaining factors accounted for only 1% to 3% each. These percentages of explained variance indicate the relative emphasis which should be given to specific procedures designed to optimize the tracking situation.

The four main effects and two interactions were significant in Exp. 2 (Table 3). The three main effects, *Ss*, Sampling Rates, and Delay-no delay, account for 90% of the explained variance. Again other statistically significant factors each account for relatively little variance.

Only Sampling Rate and the *Ss*-by-Sessions (days) interaction were significant in Exp. 3 (Table 4). Sampling rates alone, in this experiment, account for over 90% of the explained variance.

One-tailed *t* tests were conducted for the hypothesis that the continuous condition was not superior to each of the sampled conditions with optimum sensitivity in Exp. 1. The hypothesis was rejected at the .01 level of confidence or beyond in each case except the 6-cps condition, which was rejected at the .05 level. The same null hypothesis for Exp. 3 was rejected at the .01 level in each case except the 5-cps condition.

Regression curves. — Regression equations were fitted to the data from each of the experiments. Plots of mean recovery time in the first two experiments and mean 50% limits in the third experiment vs. sampling rate indicated a linear relationship on a log-log plot. Linear regression equations and correlations were calculated predicting tracking performance (log recovery time or log 50% limits) from log sampling rate.[8] These results are shown in Fig. 3, 4, and 5. These correlations, of course, indicate the goodness of fit of the regression equations. All except the correlation for Exp. 3 are significant at the .05 level or beyond.

On Exp. 1 and 3, where a continuous or nonsampled condition was studied, the mean recovery time and mean 50% limits for this continuous condition were used, with the appropriate regression equation, to predict a sampling rate which would give tracking performance equivalent to continuous sampling. These continuous-equivalent sampling rates are shown with

[8] Similar relationships between tracking performance and sampling *period* (*T*), i.e., time between samples, could be plotted. These relationships would be of the same mathematical form as those with *sampling rate* (*S*) as may be determined by substituting $T = 1/S$ in the regression equations.

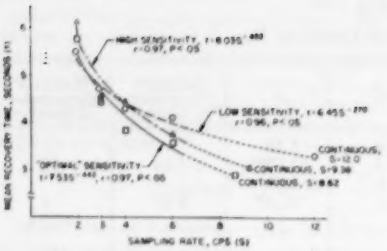

FIG. 3. Experiment 1—regression of mean recovery time on sampling rate.

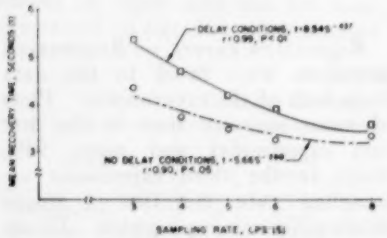

FIG. 4. Experiment 2—regression of mean recovery time on sampling rate.

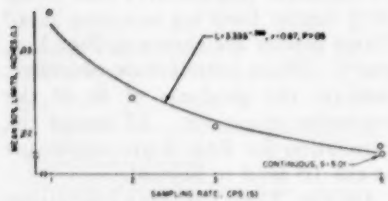

FIG. 5. Experiment 3—regression of mean 50% limits on sampling rate.

the corresponding regression curves in Fig. 3, 4, and 5.

DISCUSSION

Sampling rate.—The analysis of variance results of the three experiments consistently show the importance of sampling rate as an influence of tracking performance—the percentage of explained variance attributable to this source being 52, 33, and 92, respectively.

The regression equations serve to predict what particular rate would be equivalent to analog conditions for a given tracking situation. Thus, for a tracking situation involving targets which are discretely displaced from the cross hairs, such as studied in Exp. 1, the prediction is that a sampling rate of about 9 cps would be equivalent to continuous sampling tracking. In contrast to the rather high continuous-equivalent sampling rate predicted for the discretely displaced target situation, this predicted rate for the tracking of constantly displaced targets is 5 cps.

It is interesting to speculate concerning this difference in the predicted continuous-equivalent sampling rates for the two types of tracking situations—discretely displaced and constantly displaced targets.

In the constantly displaced tracking situation, *S* "lets" the target move slowly off the cross hairs some small distance and then attempts to return it. Once the error is nulled, a correct displacement of the hand control will produce a signal equal in magnitude and opposite in direction to the disturbance rate. Such a hand-control displacement would keep the error nulled as long as the particular disturbance is active. To achieve such a situation, however, is a rather difficult task, and generally *S* must again eventually return the target which has "crept" out some distance from the cross hairs. The tracking of such constantly displaced targets seems to consist, in part, of a series of tracking activities similar in nature to those performed with discretely displaced targets. This leaves the magnitude of the disturbances (the distance through which the target must be returned) as a possible explanation for the difference in continuous-equivalent sampling rate.

An analysis of variance of the initial slope of the response curves for the last session of Exp. 1 was carried out. Since these curves indicate target position as a function of time, the slope indicates rate of movement of the target (as produced by *S*), i.e., the utilized sensitivity for a given response. This slope is not constant, but is nearly so, initially. This analysis of variance indicates that the

magnitude of the disturbance has a significant influence on utilized sensitivity. In particular, for the 2-in. disturbances, the utilized sensitivities are nearly twice as large, on the average, as for the 1-in. disturbances. The analysis of recovery times for Exp. 1 indicates that, for higher sampling rates, higher sensitivities are optimum. Thus, it may be reasoned that, for larger disturbances, higher sensitivities and higher sampling rates are required. The discrete displacements were larger than the displacements resulting from the integration of rate of displacement obtained with the constantly displaced targets. Thus, higher sampling rates for continuous equivalence would seem to be required for the discretely displaced targets.

Delay.—In Exp. 2, the delay-no delay main effect accounts for nearly 20% of the explained variance. The delay by sampling rate interaction accounts for about 3%. The effects are in the predicted direction—delay producing considerable degradation of tracking performance, and, in particular, more degradation at the lower sampling rates where the delay is longer. This fact points to one possible explanation of the effects of sampling rate, per se, on tracking performance. As previously pointed out, sampling itself introduces delays varying from one to two sampling periods, with a mean of one and one-half sampling periods. Thus, the delay at a lower rate is longer than at a higher rate, and performance is presumably poorer.

Sensitivity.—The sensitivity main effect in Exp. 1 is not significant. This may be due to the fact that all three sensitivities ("high," "low," "optimum") were chosen to be in the "usable" range, and thus the principal effects of sensitivity were eliminated before the experiment began. The sensitivity by sampling rate interaction was significant. Examination of means for the various sampling rates and sensitivities indicates that his interaction is principally due to the superiority of the "low" sensitivity at 2 cps. Tracking under this condition was better than under "optimum" sensitivity, owing to the relatively crude method of choosing these sensitivities prior to the experiment.

Subjects.—The *S*s' main effects were significant and contributed about 40% of the sum of explained variance components in the first two experiments, but were not even significant in the third experiment. This is probably due to some difference in the task, which also permitted learning in the first two experiments (particularly in the first three trials which were not included in the analysis of variance) and no learning in the third experiment. Subjects by Sessions interaction was significant in Exp. 1 and 3, as was the Sessions main effect in Exp. 2. These effects reflect *decreased* rather than increased performance, and thus may be attributed to decreased motivation, rather than to effects of learning, per se.

Subjects by Sampling Rate was significant in Exp. 1. Study of means reveals that this was due to very poor performance by some *S*s at very low sampling rates. Subjects by Sensitivities was also significant in Exp. 1. This effect was due to the fact that one *S* performed better at the low sensitivities than at optimum sensitivities. It is difficult to ascribe these effects to any factor controlled in these experiments. In any case, it may be pointed out that in the present instance, these interactions contributed very small percentages of the total explained variance.

Two third-order interactions were significant. These interactions were: (*a*) the *S*s by Sampling Rates by Sensitivities by Sessions in Exp. 1; and (*b*) the *S*s by Sampling Rates by Delays by Sessions in Exp. 2. Interpretation of these effects appears impossible.

Summary

In the situation where an operator closes a control loop by his tracking behavior, the usual analog loop may be altered by sampling of the input to the operator or by sampling the operator's output. Three experiments, in which *S*'s output was sampled, established the importance of sampling rate as a determinant of tracking performance. A functional relationship (track-

ing performance is proportional to the sampling rate raised to some power) was established, such that tracking performance is poorer at lower sampling rates.

The variables of hand-control sensitivity and operator response-effect delays were found to affect sampled-data tracking performance.

References

1. Battig, W. F., Gregg, L. W., Nagel, E. H., Small, A. M., Jr., & Brogden, W. J. Tracking and frequency of target intermittence. *J. exp. Psychol.*, 1954, **47**, 309–314.
2. Battig, W. F., Voss, J. F., & Brogden, W. J. Effect of frequency of target intermittence upon tracking. *J. exp. Psychol.*, 1955, **49**, 244–248.
3. Chapanis, A., Garner, W. R., & Morgan, C. T. *Applied experimental psychology.* New York: Wiley, 1949.
4. Katz, M. S., & Spragg, S. D. S. Tracking performance as a function of frequency of course illumination. *J. Psychol.*, 1955, **40**, 181–192.
5. Mood, A. F. *Introduction to a theory of statistics.* New York: McGraw-Hill, 1950.
6. Rockway, M. R. The effect of variations in control-display ratio and exponential time delay on tracking performance. *USAF, WADC Tech. Rep.*, 1954, No. 54-618.
7. Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.
8. Warrick, M. J. Effect of transmission-type control logs on tracking accuracy. *USAF, WADC Tech. Rep.*, 1949, No. 5916.

(Received for early publication November 17, 1955)